

Continuation de 337 800

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/85, 9/10, C07K 16/30, 16/40, C12Q 1/68, A01K 67/027, C07K 14/47		A3	(11) International Publication Number: WO 99/32644
			(43) International Publication Date: 1 July 1999 (01.07.99)
(21) International Application Number: PCT/IB98/02133		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: 22 December 1998 (22.12.98)		Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i> (88) Date of publication of the international search report: 10 September 1999 (10.09.99)	
(30) Priority Data: 08/996,306 22 December 1997 (22.12.97) US 60/099,658 9 September 1998 (09.09.98) US			
(71) Applicant (for all designated States except US): GENSET [FR/FR]; 24, rue Royale, F-75008 Paris (FR).			
(72) Inventors; and (75) Inventors/Applicants (for US only): COHEN, Daniel [FR/FR]; 5, avenue Odette, F-94210 Fontenay-sous-Bois (FR). BLUMENFELD, Marta [FR/FR]; 5, rue Tagore, F-75013 Paris (FR). CHUMAKOV, Ilya [FR/FR]; 196, rue des Chèvrefeuilles, F-77000 Vaux-le-Pénil (FR). BOUGUEL-ERET, Lydie [FR/FR]; 108, avenue Victor-Hugo, F-92170 Vanves (FR).			
(74) Agents: MARTIN, Jean-Jacques et al.; Cabinet Regimbeau, 26, avenue Kléber, F-75116 Paris (FR).			
(54) Title: PROSTATE CANCER GENE			
(57) Abstract			
<p>The present invention relates to PG1, a gene associated with prostate cancer. The invention provides polynucleotides including biallelic markers derived from PG1 and from flanking genomic regions. Primers hybridizing to these biallelic markers and regions flanking are also provided. This invention provides polynucleotides and methods suitable for genotyping a nucleic acid containing sample for one or more biallelic markers of the invention. Further, the invention provides methods to detect a statistical correlation between a biallelic marker allele and prostate cancer and between a haplotype and prostate cancer. The invention also relates to diagnostic methods of determining whether an individual is at risk for developing prostate cancer, and whether an individual suffers from prostate cancer as a result of a mutation in the PG1 gene.</p>			

INTERNATIONAL SEARCH REPORT

Inte. .onal Application No

PCT/IB 98/02133

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/85 C12N9/10 C07K16/30 C07K16/40 C12Q1/68
A01K67/027 C07K14/47

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12Q C07K C12N A01K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EMBL Database Entry HS164152 Accession Number H06164; 23 June 1995 Hillier L et al.: "y177g12.r1 Homosapiens cDNA clone 44264 5'." 100% identity in 169pb overlap with Seq ID No 118. XP002109141 see database entry --- -/--	1-3



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"Δ" document member of the same patent family

Date of the actual completion of the international search

15 July 1999

Date of mailing of the international search report

29/07/1999

Name and mailing address of the ISA

European Patent Office, P B 5618 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Osborne, H

INTERNATIONAL SEARCH REPORT

International Application No
PCT/IB 98/02133

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>KRUGLYAK L: "THE USE OF A GENETIC MAP OF BIALLELIC MARKERS IN LINKAGE STUDIES" NATURE GENETICS, vol. 17, no. 1, 1 September 1997 (1997-09-01), pages 22-24, XP002050647 ISSN: 1061-4036 the whole document ---</p>	19-38
A	<p>WANG D ET AL: "TOWARD A THIRD GENERATION GENETIC MAP OF THE HUMAN GENOME BASED ON BI-ALLELIC POLYMORPHISMS" AMERICAN JOURNAL OF HUMAN GENETICS, vol. 59, no. 4, 1 October 1996 (1996-10-01), page A03 XP002050641 ISSN: 0002-9297 the whole document ---</p>	19-38
A	<p>SCHORK N J ET AL: "Linkage disequilibrium mapping for quantitative traits within case/control settings" AMERICAN JOURNAL OF HUMAN GENETICS, vol. 61, no. 4, SUPPL, 1 January 1997 (1997-01-01), page A293 XP002089399 ISSN: 0002-9297 -----</p>	19-38

INTERNATIONAL SEARCH REPORT

International Application No. PCT/IB 98 02133

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 4 and 39

Present claims 4 and 39 relate to an extremely large number of possible polynucleotide sequences. In the present case, a meaningful search over the whole of the claimed scope is impossible. Consequently, the search has been carried out for those parts of the claims which appear to be related to the PGI gene per se., and their utility in the claimed diagnostic methods.

The applicant's attention is drawn to the fact that claims relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶: C12N 15/85, 9/10, C07K 16/30, 16/40, C12Q 1/68, A01K 67/027	A2	(11) International Publication Number: WO 99/32644 (43) International Publication Date: 1 July 1999 (01.07.99)
(21) International Application Number: PCT/IB98/02133 (22) International Filing Date: 22 December 1998 (22.12.98) (30) Priority Data: 08/996,306 22 December 1997 (22.12.97) US 60/099,658 9 September 1998 (09.09.98) US (71) Applicant (for all designated States except US): GENSET [FR/FR]; 24, rue Royale, F-75008 Paris (FR). (72) Inventors; and (75) Inventors/Applicants (for US only): COHEN, Daniel [FR/FR]; 5, avenue Odette, F-94210 Fontenay-sous-Bois (FR). BLUMENFELD, Marta [FR/FR]; 5, rue Tagore, F-75013 Paris (FR). CHUMAKOV, Ilya [FR/FR]; 196, rue des Chèvrefeuilles, F-77000 Vaux-le-Pénil (FR). BOUGUEL- ERET, Lydie [FR/FR]; 108, avenue Victor-Hugo, F-92170 Vanves (FR). (74) Agents: MARTIN, Jean-Jacques et al.; Cabinet Regimbeau, 26, avenue Kléber, F-75116 Paris (FR).		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published Without international search report and to be republished upon receipt of that report.
(54) Title: PROSTATE CANCER GENE (57) Abstract <p>The present invention relates to PG1, a gene associated with prostate cancer. The invention provides polynucleotides including biallelic markers derived from PG1 and from flanking genomic regions. Primers hybridizing to these biallelic markers and regions flanking are also provided. This invention provides polynucleotides and methods suitable for genotyping a nucleic acid containing sample for one or more biallelic markers of the invention. Further, the invention provides methods to detect a statistical correlation between a biallelic marker allele and prostate cancer and between a haplotype and prostate cancer. The invention also relates to diagnostic methods of determining whether an individual is at risk for developing prostate cancer, and whether an individual suffers from prostate cancer as a result of a mutation in the PG1 gene.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

PROSTATE CANCER GENE

Background of the Invention

5 A cancer is a clonal proliferation of cells produced as a consequence of cumulative genetic damage that finally results in unrestrained cell growth, tissue invasion and metastasis (cell transformation). Regardless of the type of cancer, transformed cells carry damaged DNA in many forms: as gross chromosomal translocations or, more subtly, as DNA amplification, rearrangement or even point mutations.

10 Some oncogenic mutations is inherited in the germline, thus predisposing the mutation carrier to an increased risk of cancer. However, in a majority of cases, cancer does not occur as a simple monogenic disease with clear Mendelian inheritance. There is only a two- or threefold increased risk of cancer among first-degree relatives for many cancers (Mulvihill JJ, Miller RW & Fraumeni JF, 1977, Genetics of human cancer Vol 3, New York Raven Press). Alternatively, DNA damage is
15 acquired somatically, probably induced by exposure to environmental carcinogens. Somatic mutations are generally responsible for the vast majority of cancer cases.

Studies of the age dependence of cancer have suggested that several successive mutations are needed to convert a normal cell into an invasive carcinoma. Since human mutation rates are typically 10^{-6} /gene/cell, the chance of a single cell undergoing many independent mutations is very low (Loeb
20 LA, Cancer Res 1991, 51: 3075-3079). Cancer nevertheless happens because of a combination of two mechanisms. Some mutations enhance cell proliferation, increasing the target population of cells for the next mutation. Other mutations affect the stability of the entire genome, increasing the overall mutation rate, as in the case of mismatch repair proteins (reviewed in Arnheim N & Shibata D, Curr. Op. Genetics & Development, 1997, 7:364-370).

25 An intricate process known as the cell cycle drives normal proliferation of cells in an organism. Regulation of the extent of cell cycle activity and the orderly execution of sequential steps within the cycle ensure the normal development and homeostasis of the organism. Conversely, many of the properties of cancer cells - uncontrolled proliferation, increased mutation rate, abnormal translocations and gene amplifications - can be attributed directly to perturbations of the normal
30 regulation or progression of the cycle. In fact, many of the genes that have been identified over the past several decades as being involved in cancer, can now be appreciated in terms of their direct or indirect role in either regulating entry into the cell cycle or coordinating events within the cell cycle.

Recent studies have identified three groups of genes which are frequently mutated in cancer. The first group of genes, called oncogenes, are genes whose products activate cell proliferation. The
35 normal non-mutant versions are called protooncogenes. The mutated forms are excessively or inappropriately active in promoting cell proliferation, and act in the cell in a dominant way in that a

single mutant allele is enough to affect the cell phenotype. Activated oncogenes are rarely transmitted as germline mutations since they may probably be lethal when expressed in all the cells. Therefore oncogenes can only be investigated in tumor tissues.

Oncogenes and protooncogenes can be classified into several different categories according to their function. This classification includes genes that code for proteins involved in signal transduction such as: growth factors (i.e., sis, int-2); receptor and non-receptor protein-tyrosine kinases (i.e., erbB, src, bcr-abl, met, trk); membrane-associated G proteins (i.e., ras); cytoplasmic protein kinases (i.e., mitogen-activated protein kinase -MAPK- family, raf, mos, pak), or nuclear transcription factors (i.e., myc, myb, fos, jun, rel) (for review see Hunter T, 1991 Cell 64:249; Fanger GR et al., 1997 Curr.Op.Genet.Dev.7:67-74; Weiss FU et al., ibid. 80-86).

The second group of genes which are frequently mutated in cancer, called tumor suppressor genes, are genes whose products inhibit cell growth. Mutant versions in cancer cells have lost their normal function, and act in the cell in a recessive way in that both copies of the gene must be inactivated in order to change the cell phenotype. Most importantly, the tumor phenotype can be rescued by the wild type allele, as shown by cell fusion experiments first described by Harris and colleagues (Harris H et al., 1969, Nature 223:363-368). Germline mutations of tumor suppressor genes is transmitted and thus studied in both constitutional and tumor DNA from familial or sporadic cases. The current family of tumor suppressors includes DNA-binding transcription factors (i.e., p53, WT1), transcription regulators (i.e., RB, APC, probably BRCA1), protein kinase inhibitors (i.e., p16), among others (for review, see Haber D & Harlow E, 1997, Nature Genet. 16:320-322).

The third group of genes which are frequently mutated in cancer, called mutator genes, are responsible for maintaining genome integrity and/or low mutation rates. Loss of function of both alleles increase cell mutation rates, and as consequence, proto-oncogenes and tumor suppressor genes is mutated. Mutator genes can also be classified as tumor suppressor genes, except for the fact that tumorigenesis caused by this class of genes cannot be suppressed simply by restoration of a wild-type allele, as described above. Genes whose inactivation may lead to a mutator phenotype include mismatch repair genes (i.e., MLH1, MSH2), DNA helicases (i.e., BLM, WRN) or other genes involved in DNA repair and genomic stability (i.e., p53, possibly BRCA1 and BRCA2) (For review see Haber D & Harlow E, 1997, Nature Genet. 16:320-322; Fishel R & Wilson T. 1997, Curr.Op.Genet.Dev.7: 105-113; Ellis NA, 1997 ibid.354-363).

The recent development of sophisticated techniques for genetic mapping has resulted in an ever expanding list of genes associated with particular types of human cancers. The human haploid genome contains an estimated 80,000 to 100,000 genes scattered on a 3×10^9 base-long double-stranded DNA. Each human being is diploid, i.e., possesses two haploid genomes, one from paternal origin, the other from maternal origin. The sequence of a given genetic locus may vary between individuals in a population or between the two copies of the locus on the chromosomes of a single

individual. Genetic mapping techniques often exploit these differences, which are called polymorphisms, to map the location of genes associated with human phenotypes.

One mapping technique, called the loss of heterozygosity (LOH) technique, is often employed to detect genes in which a loss of function results in a cancer, such as the tumor suppressor genes described above. Tumor suppressor genes often produce cancer via a two hit mechanism in which a first mutation, such as a point mutation (or a small deletion or insertion) inactivates one allele of the tumor suppressor gene. Often, this first mutation is inherited from generation to generation.

A second mutation, often a spontaneous somatic mutation such as a deletion which deletes all or part of the chromosome carrying the other copy of the tumor suppressor gene, results in a cell in which both copies of the tumor suppressor gene are inactive.

As a consequence of the deletion in the tumor suppressor gene, one allele is lost for any genetic marker located close to the tumor suppressor gene. Thus, if the patient is heterozygous for a marker, the tumor tissue loses heterozygosity, becoming homozygous or hemizygous. This loss of heterozygosity generally provides strong evidence for the existence of a tumor suppressor gene in the lost region.

By genotyping pairs of blood and tumor samples from affected individuals with a set of highly polymorphic genetic markers, such as microsatellites, covering the whole genome, one can discover candidate locations for tumor suppressor genes. Due to the presence of contaminant non-tumor tissue in most pathological tumor samples, a decreased relative intensity rather than total loss of heterozygosity of informative microsatellites is observed in the tumor samples. Therefore, classic LOH analysis generally requires quantitative PCR analysis, often limiting the power of detection of this technique. Another limitation of LOH studies resides on the fact that they only allow the definition of rather large candidate regions, typically spanning over several megabases. Refinement of such candidate regions requires the definition of the minimally overlapping portion of LOH regions identified in tumor tissues from several hundreds of affected patients.

Another approach to genetic mapping, called linkage analysis, is based upon establishing a correlation between the transmission of genetic markers and that of a specific trait throughout generations within a family. In this approach, all members of a series of affected families are genotyped with a few hundred markers, typically microsatellite markers, which are distributed at an average density of one every 10 Mb. By comparing genotypes in all family members, one can attribute sets of alleles to parental haploid genomes (haplotyping or phase determination). The origin of recombined fragments is then determined in the offspring of all families. Those that co-segregate with the trait are tracked. After pooling data from all families, statistical methods are used to determine the likelihood that the marker and the trait are segregating independently in all families. As a result of the statistical analysis, one or several regions are selected as candidates, based on their high probability to carry a trait causing allele. The result of linkage analysis is considered as significant

when the chance of independent segregation is lower than 1 in 1000 (expressed as a LOD score > 3). Identification of recombinant individuals using additional markers allows further delineation of the candidate linked region, which most usually ranges from 2 to 20 Mb.

Linkage analysis studies have generally relied on the use of microsatellite markers (also
5 called simple tandem repeat polymorphisms, or simple sequence length polymorphisms). These include small arrays of tandem repeats of simple sequences (di- tri- tetra- nucleotide repeats), which exhibit a high degree of length polymorphism, and thus a high level of informativeness. To date, only
just more than 5,000 microsatellites have been ordered along the human genome (Dib et al., Nature
1996, 380: 152), thus limiting the maximum attainable resolution of linkage analysis to ca. 600 kb on
10 average.

Linkage analysis has been successfully applied to map simple genetic traits that show clear Mendelian inheritance patterns. About 100 pathological trait-causing genes were discovered by linkage analysis over the last 10 years.

However, linkage analysis approaches have proven difficult for complex genetic traits, those
15 probably due to the combined action of multiple genes and/or environmental factors. In such cases, too large an effort and cost are needed to recruit the adequate number of affected families required for applying linkage analysis to these situations, as recently discussed by Risch, N. and Merikangas, K. (Science 1996, 273: 1516-1517). Finally, linkage analysis cannot be applied to the study of traits for which no available large informative families are available. Typically, this will be the case in any
20 attempt to identify trait-causing alleles involved in sporadic cases.

The incidence of prostate cancer has dramatically increased over the last decades. It averages 30-50/100,000 males both in Western European countries as well as within the US White male population. In these countries, it has recently become the most commonly diagnosed malignancy, being one of every four cancers diagnosed in American males. Prostate cancer's incidence is very
25 much population specific, since it varies from 2/100,000 in China, to over 80/100,000 among African-American males.

In France, the incidence of prostate cancer is 35/100,000 males and it is increasing by 10/100,000 per decade. Mortality due to prostate cancer is also growing accordingly. It is the second cause of cancer death among French males, and the first one among French males aged over 70. This
30 makes prostate cancer a serious burden in terms of public health, especially in view of the aging of populations.

An average 40% reduction in life expectancy affects males with prostate cancer. If completely localized, prostate cancer can be cured by surgery, with however an average success rate of only ca. 50%. If diagnosed after metastasis from the prostate, prostate cancer is a fatal disease for
35 which there is no curative treatment.

Early-stage diagnosis relies on Prostate Specific Antigen (PSA) dosage, and would allow the

detection of prostate cancer seven years before clinical symptoms become apparent. The effectiveness of PSA dosage diagnosis is however limited, due to its inability to discriminate between malignant and non-malignant affections of the organ.

Therefore, there is a strong need for both a reliable diagnostic procedure which would enable
5 early-stage prostate cancer prognosis, and for preventive and curative treatments of the disease. The present invention relates to the PG1 gene, a gene associated with prostate cancer, as well as diagnostic methods and reagents for detecting alleles of the gene which may cause prostate cancer, and therapies for treating prostate cancer.

10 Summary of the Invention

The present invention relates to the identification of a gene associated with prostate cancer, identified as the PG1 gene, and reagents, diagnostics, and therapies related thereto. The present invention is also based on the discovery of a novel set of PG1-related biallelic markers. See the definition of PG1-related biallelic markers in the Detailed Description Section. These markers are
15 located in the coding regions as well as non-coding regions adjacent to the PG1 gene. The position of these markers and knowledge of the surrounding sequence has been used to design polynucleotide compositions which are useful in determining the identity of nucleotides at the marker position, as well as more complex association and haplotyping studies which are useful in determining the genetic basis for diseases including cancer and prostate cancer. In addition, the compositions and methods of
20 the invention find use in the identification of the targets for the development of pharmaceutical agents and diagnostic methods, as well as the characterization of the differential efficacious responses to and side effects from pharmaceutical agents acting on diseases including cancer and prostate cancer.

A first embodiment of the invention is a recombinant, purified or isolated polynucleotide comprising, or consisting of a mammalian genomic sequence, gene, or fragments thereof. In one
25 aspect the sequence is derived from a human, mouse or other mammal. In a preferred aspect, the genomic sequence is the human genomic sequence of SEQ ID NO: 179 or the complement thereto. In a second preferred aspect, the genomic sequence is selected from one of the two mouse genomic fragments of SEQ ID NO: 182 and 183. In yet another aspect of this embodiment, the nucleic acid comprises nucleotides 1629 through 1870 of the sequence of SEQ ID NO: 179. Optionally, said
30 polynucleotide consists of, consists essentially of, or comprises a contiguous span of nucleotides of a mammalian genomic sequence, preferably a sequence selected the following SEQ ID NOs: 179, 182, and 183, wherein said contiguous span is at least 6, 8, 10, 12, 15, 20, 25, 30, 50, 100, 200, or 500 nucleotides in length.

A second embodiment of the present invention is a recombinant, purified or isolated
35 polynucleotide comprising, or consisting of a mammalian cDNA sequence, or fragments thereof. In one aspect the sequence is derived from a human, mouse or other mammal. In a preferred aspect, the

cDNA sequence is selected from the human cDNA sequences of SEQ ID NO: 3, 69, 112-124 or the complement thereto. In a second preferred aspect, the cDNA sequence is the mouse cDNA sequence of SEQ ID NO: 184. Optionally, said polynucleotide consists of, consists essentially of, or comprises a contiguous span of nucleotides of a mammalian genomic sequence, preferably a sequence selected
5 the following SEQ ID NOs: 3, 69, 112-124 and 184, wherein said contiguous span is at least 6, 8, 10, 12, 15, 20, 25, 30, 50, 100, 200, or 500 nucleotides in length.

A third embodiment of the present invention is a recombinant, purified or isolated polynucleotide, or the complement thereof, encoding a mammalian PG1 protein, or a fragment thereof. In one aspect the PG1 protein sequence is from a human, mouse or other mammal. In a
10 preferred aspect, the PG1 protein sequence is selected from the human PG1 protein sequences of SEQ ID NO: 4, 5, 70, and 125-136. In a second preferred aspect, the PG1 protein sequence is the mouse PG1 protein sequences of SEQ ID NO: 74. Optionally, said fragment of PG1 polypeptide consists of, consists essentially of, or comprises a contiguous stretch of at least 8, 10, 12, 15, 20, 25, 30, 50, 100 or 200 amino acids from SEQ ID NOs: 4, 5, 70, 74, and 125-136, as well as any other human, mouse
15 or mammalian PG1 polypeptide.

A fourth embodiment of the invention are the polynucleotide primers and probes disclosed herein

A fifth embodiment of the present invention is a recombinant, purified or isolated polypeptide comprising or consisting of a mammalian PG1 protein, or a fragment thereof. In one aspect the PG1
20 protein sequence is from a human, mouse or other mammal. In a preferred aspect, the PG1 protein sequence is selected from the human PG1 protein sequences of SEQ ID NO: 4, 5, 70, and 125-136. In a second preferred aspect, the PG1 protein sequence is the mouse PG1 protein sequences of SEQ ID NO: 74. Optionally, said fragment of PG1 polypeptide consists of, consists essentially of, or comprises a contiguous stretch of at least 8, 10, 12, 15, 20, 25, 30, 50, 100 or 200 amino acids from
25 SEQ ID NOs: 4, 5, 70, 74, and 125-136, as well as any other human, mouse or mammalian PG1 polypeptide.

A sixth embodiment of the present invention is an antibody composition capable of specifically binding to a polypeptide of the invention. Optionally, said antibody is polyclonal or monoclonal. Optionally, said polypeptide is an epitope-containing fragment of at least 8, 10, 12, 15,
30 20, 25, or 30 amino acids of a human, mouse, or mammalian PG1 protein, preferably a sequence selected from SEQ ID NOs: 4, 5, 70, 74, or 125-136.

A seventh embodiment of the present invention is a vector comprising any polynucleotide of the invention. Optionally, said vector is an expression vector, gene therapy vector, amplification vector, gene targeting vector, or knock-out vector.

35 An eighth embodiment of the present invention is a host cell comprising any vector of the invention.

A ninth embodiment of the present invention is a mammalian host cell comprising a PG1 gene disrupted by homologous recombination with a knock out vector.

A tenth embodiment of the present invention is a nonhuman host mammal or animal comprising a vector of the invention.

5 A further embodiment of the present invention is a nonhuman host mammal comprising a PG1 gene disrupted by homologous recombination with a knock out vector.

Another embodiment of the present invention is a method of determining whether an individual is at risk of developing cancer or prostate cancer at a later date or whether the individual suffers from cancer or prostate cancer as a result of a mutation in the PG1 gene comprising obtaining
10 a nucleic acid sample from the individual; and determining whether the nucleotides present at one or more of the PG1-related biallelic markers of the invention are indicative of a risk of developing prostate cancer at a later date or indicative of prostate cancer resulting from a mutation in the PG1 gene. Optionally, said PG1-related biallelic is a PG1-related biallelic markers positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected from the group consisting of 99-1485/251, 99-
15 622/95, 99-619/141, 4-76/222, 4-77/151, 4-71/233, 4-72/127, 4-73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-600/492, 99-598/130, 99-217/277, 99-576/421, 4-61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576, and 4-66.

Another embodiment of the present invention is a method of determining whether an
20 individual is at risk of developing prostate cancer at a later date or whether the individual suffers from prostate cancer as a result of a mutation in the PG1 gene comprising obtaining a nucleic acid sample from the individual and determining whether the nucleotides present at one or more of the polymorphic bases in a PG1-related biallelic marker. Optionally, said PG1-related biallelic is a PG1-related biallelic markers positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected from
25 the group consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-77/151, 4-71/233, 4-72/127, 4-73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-600/492, 99-598/130, 99-217/277, 99-576/421, 4-61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576, and 4-66.

Another embodiment of the present invention is a method of obtaining an allele of the PG1
30 gene which is associated with a detectable phenotype comprising obtaining a nucleic acid sample from an individual expressing the detectable phenotype, contacting the nucleic acid sample with an agent capable of specifically detecting a nucleic acid encoding the PG1 protein, and isolating the nucleic acid encoding the PG1 protein. In one aspect of this method, the contacting step comprises contacting the nucleic acid sample with at least one nucleic acid probe capable of specifically
35 hybridizing to said nucleic acid encoding the PG1 protein. In another aspect of this embodiment, the contacting step comprises contacting the nucleic acid sample with an antibody capable of specifically

binding to the PG1 protein. In another aspect of this embodiment, the step of obtaining a nucleic acid sample from an individual expressing a detectable phenotype comprises obtaining a nucleic acid sample from an individual suffering from prostate cancer.

Another embodiment of the present invention is a method of obtaining an allele of the PG1
5 gene which is associated with a detectable phenotype comprising obtaining a nucleic acid sample from an individual expressing the detectable phenotype, contacting the nucleic acid sample with an agent capable of specifically detecting a sequence within the 8p23 region of the human genome, identifying a nucleic acid encoding the PG1 protein in the nucleic acid sample, and isolating the nucleic acid encoding the PG1 protein. In one aspect of this embodiment, the nucleic acid sample is
10 obtained from an individual suffering from cancer or prostate cancer.

Another embodiment of the present invention is a method of categorizing the risk of prostate cancer in an individual comprising the step of assaying a sample taken from the individual to determine whether the individual carries an allelic variant of PG1 associated with an increased risk of prostate cancer. In one aspect of this embodiment, the sample is a nucleic acid sample. In another
15 aspect a nucleic acid sample is assayed by determining the frequency of the PG1 transcripts present. In another aspect of this embodiment, the sample is a protein sample. In another aspect of this embodiment, the method further comprises determining whether the PG1 protein in the sample binds an antibody specific for a PG1 isoform associated with prostate cancer.

Another embodiment of the present invention is a method of categorizing the risk of prostate
20 cancer in an individual comprising the step of determining whether the identities of the polymorphic bases of one or more biallelic markers which are in linkage disequilibrium with the PG1 gene are indicative of an increased risk of prostate cancer.

Another embodiment of the present invention comprises a method of identifying molecules which specifically bind to a PG1 protein, preferably the protein of SEQ ID NO:4 or a portion thereof:
25 comprising the steps of introducing a nucleic acid encoding the protein of SEQ ID NO:4 or a portion thereof into a cell such that the protein of SEQ ID NO:4 or a portion thereof contacts proteins expressed in the cell and identifying those proteins expressed in the cell which specifically interact with the protein of SEQ ID NO:4 or a portion thereof.

Another embodiment of the present invention is a method of identifying molecules which
30 specifically bind to the protein of SEQ ID NO: 4 or a portion thereof. One step of the method comprises linking a first nucleic acid encoding the protein of SEQ ID NO:4 or a portion thereof to a first indicator nucleic acid encoding a first indicator polypeptide to generate a first chimeric nucleic acid encoding a first fusion protein. The first fusion protein comprises the protein of SEQ ID NO:4 or a portion thereof and the first indicator polypeptide. Another step of the method comprises linking a
35 second nucleic acid encoding a test polypeptide to a second indicator nucleic acid encoding a second indicator polypeptide to generate a second chimeric nucleic acid encoding a second

fusion protein. The second fusion protein comprises the test polypeptide and the second indicator polypeptide. Association between the first indicator protein and the second indicator protein produces a detectable result. Another step of the method comprises introducing the first chimeric nucleic acid and the second chimeric nucleic acid into a cell. Another step comprises detecting the detectable
5 result.

A further embodiment of the invention is a purified or isolated mammalian PG1 gene or cDNA sequence.

Further embodiments of the present invention include the nucleic acid and amino acid sequences of mutant or low frequency PG1 alleles derived from prostate cancer patients, tissues or
10 cell lines. The present invention also encompasses methods which utilize detection of these mutant PG1 sequences in an individual or tissue sample to diagnosis prostate cancer, assess the risk of developing prostate cancer or assess the likely severity of a particular prostate tumor.

Another embodiment of the invention encompasses any polynucleotide of the invention attached to a solid support. In addition, the polynucleotides of the invention which are attached to a
15 solid support encompass polynucleotides with any further limitation described in this disclosure, or those following: Optionally, said polynucleotides is specified as attached individually or in groups of at least 2, 5, 8, 10, 12, 15, 20, or 25 distinct polynucleotides of the inventions to a single solid support. Optionally, polynucleotides other than those of the invention may attached to the same solid support as polynucleotides of the invention. Optionally, when multiple polynucleotides are attached to a solid
20 support they are attached at random locations, or in an ordered array. Optionally, said ordered array is addressable.

An additional embodiment of the invention encompasses the use of any polynucleotide for, or any polynucleotide for use in, determining the identity of an allele at a PG1-related biallelic marker. In addition, the polynucleotides of the invention for use in determining the identity of an allele at a
25 PG1-related biallelic marker encompass polynucleotides with any further limitation described in this disclosure, or those following: Optionally, said PG1-related biallelic marker is a PG1-related biallelic markers positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected from the group consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-77/151, 4-71/233, 4-72/127, 4-73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-600/492, 99-598/130, 99-217/277, 99-576/421, 4-
30 61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576, and 4-66. Optionally, said polynucleotide may comprise a sequence disclosed in the present specification. Optionally, said polynucleotide may consist of, or consist essentially of any polynucleotide described in the present specification. Optionally, said determining is performed in a hybridization assay, sequencing assay, microsequencing assay, or allele-specific amplification assay. Optionally, said polynucleotide is
35 attached to a solid support, array, or addressable array. Optionally, said polynucleotide is labeled.

Another embodiment of the invention encompasses the use of any polynucleotide for, or any polynucleotide for use in, amplifying a segment of nucleotides comprising an PG1-related biallelic marker. In addition, the polynucleotides of the invention for use in amplifying a segment of nucleotides comprising a PG1-related biallelic marker encompass polynucleotides with any further
5 limitation described in this disclosure, or those following: Optionally, said PG1-related biallelic marker is a PG1-related biallelic markers positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected from the group consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-77/151, 4-71/233, 4-72/127, 4-73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-600/492, 99-598/130, 99-217/277, 99-576/421, 4-61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic
10 marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576, and 4-66. Optionally, said polynucleotide may comprise a sequence disclosed in the present specification. Optionally, said polynucleotide may consist of, or consist essentially of any polynucleotide described in the present specification. Optionally, said amplifying is performed by a PCR or LCR. Optionally, said polynucleotide is attached to a solid support, array, or addressable array. Optionally, said
15 polynucleotide is labeled.

A further embodiment of the invention encompasses methods of genotyping a biological sample comprising determining the identity of an allele at an PG1-related biallelic marker. In addition, the genotyping methods of the invention encompass methods with any further limitation described in this disclosure, or those following: Optionally, said PG1-related biallelic marker is a PG1-related
20 biallelic markers positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected from the group consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-77/151, 4-71/233, 4-72/127, 4-73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-600/492, 99-598/130, 99-217/277, 99-576/421, 4-61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576, and 4-66. Optionally, said method further
25 comprises determining the identity of a second allele at said biallelic marker, wherein said first allele and second allele are not base paired (by Watson & Crick base pairing) to one another. Optionally, said biological sample is derived from a single individual or subject. Optionally, said method is performed in vitro. Optionally, said biallelic marker is determined for both copies of said biallelic marker present in said individual's genome. Optionally, said biological sample is derived from
30 multiple subjects or individuals. Optionally, said method further comprises amplifying a portion of said sequence comprising the biallelic marker prior to said determining step. Optionally, wherein said amplifying is performed by PCR, LCR, or replication of a recombinant vector comprising an origin of replication and said portion in a host cell. Optionally, wherein said determining is performed by a hybridization assay, sequencing assay, microsequencing assay, or allele-specific amplification assay.
35 An additional embodiment of the invention comprises methods of estimating the frequency of an allele in a population comprising determining the proportional representation of an allele at a PG1-

related biallelic marker in said population. In addition, the methods of estimating the frequency of an allele in a population of the invention encompass methods with any further limitation described in this disclosure, or those following: Optionally, said PG1-related biallelic marker is a PG1-related biallelic markers positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected from the group
5 consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-77/151, 4-71/233, 4-72/127, 4-73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-600/492, 99-598/130, 99-217/277, 99-576/421, 4-61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576, and 4-66. Optionally, determining the proportional representation of an allele at a PG1-related biallelic marker is accomplished by determining the
10 identity of the alleles for both copies of said biallelic marker present in the genome of each individual in said population and calculating the proportional representation of said allele at said PG1-related biallelic marker for the population. Optionally, determining the proportional representation is accomplished by performing a genotyping method of the invention on a pooled biological sample derived from a representative number of individuals, or each individual, in said population, and
15 calculating the proportional amount of said nucleotide compared with the total.

A further embodiment of the invention comprises methods of detecting an association between a genotype and a phenotype, comprising the steps of a) genotyping at least one PG1-related biallelic marker in a trait positive population according to a genotyping method of the invention; b) genotyping said PG1-related biallelic marker in a control population according to a genotyping
20 method of the invention; and c) determining whether a statistically significant association exists between said genotype and said phenotype. In addition, the methods of detecting an association between a genotype and a phenotype of the invention encompass methods with any further limitation described in this disclosure, or those following: Optionally, said PG1-related biallelic marker is a PG1-related biallelic markers positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected
25 from the group consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-77/151, 4-71/233, 4-72/127, 4-73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-600/492, 99-598/130, 99-217/277, 99-576/421, 4-61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576, and 4-66. Optionally, said control population is a trait negative population, or a random population. Optionally, each of said
30 genotyping steps a) and b) is performed on a single pooled biological sample derived from each of said populations. Optionally, each of said genotyping of steps a) and b) is performed separately on biological samples derived from each individual in said population or a subsample thereof. Optionally, said phenotype is a disease, cancer or prostate cancer; a response to an anti-cancer agent or an anti-prostate cancer agent; or a side effect to an anti-cancer or anti-prostate cancer agent.
35 Optionally, said method comprises the additional steps of determining the phenotype in said trait positive and said control populations prior to step c).

An additional embodiment of the present invention encompasses methods of estimating the frequency of a haplotype for a set of biallelic markers in a population, comprising the steps of: a) genotyping at least one PG1-related biallelic marker for both copies of said set of biallelic marker present in the genome of each individual in said population or a subsample thereof, according to a
5 genotyping method of the invention; b) genotyping a second biallelic marker by determining the identity of the allele at said second biallelic marker for both copies of said second biallelic marker present in the genome of each individual in said population or said subsample, according to a genotyping method of the invention; and c) applying a haplotype determination method to the identities of the nucleotides determined in steps a) and b) to obtain an estimate of said frequency. In
10 addition, the methods of estimating the frequency of a haplotype of the invention encompass methods with any further limitation described in this disclosure, or those following: Optionally, said PG1-related biallelic marker is a PG1-related biallelic markers positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected from the group consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-77/151, 4-71/233, 4-72/127, 4-73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-
15 600/492, 99-598/130, 99-217/277, 99-576/421, 4-61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576, and 4-66. Optionally, said second biallelic marker is a PG1-related biallelic marker; a PG1-related biallelic markers positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected from the group consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-77/151, 4-71/233, 4-72/127, 4-
20 73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-600/492, 99-598/130, 99-217/277, 99-576/421, 4-61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576, and 4-66. Optionally, said PG1-related biallelic marker and said second biallelic marker are 4-77/151 and 4-66/145. Optionally, said haplotype determination method is an expectation-maximization algorithm.

25 An additional embodiment of the present invention encompasses methods of detecting an association between a haplotype and a phenotype, comprising the steps of: a) estimating the frequency of at least one haplotype in a trait positive population, according to a method of the invention for estimating the frequency of a haplotype; b) estimating the frequency of said haplotype in a control population, according to a method of the invention for estimating the frequency of a
30 haplotype; and c) determining whether a statistically significant association exists between said haplotype and said phenotype. In addition, the methods of detecting an association between a haplotype and a phenotype of the invention encompass methods with any further limitation described in this disclosure, or those following: Optionally, said PG1-related biallelic is a PG1-related biallelic markers positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected from the group
35 consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-77/151, 4-71/233, 4-72/127, 4-73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-600/492, 99-598/130, 99-217/277, 99-576/421, 4-

61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576, and 4-66. Optionally, said PG1-related biallelic marker and said second biallelic marker are 4-77/151 and 4-66/145. Optionally, said haplotype exhibits a p-value of $< 1 \times 10^{-3}$ in an association with a trait positive population with cancer, preferably prostate cancer. Optionally, said control population is a trait negative population, or a random population. Optionally, said phenotype is a disease, cancer or prostate cancer; a response to an anti-cancer agent or an anti-prostate cancer agent, or a side effects to an anti-cancer or anti-prostate cancer agent. Optionally, said method comprises the additional steps of determining the phenotype in said trait positive and said control populations prior to step c).

10 Additional embodiments and aspects of the present invention are set forth in the Detailed Description of the Invention and the Examples.

Brief Description of the Drawings

Figure 1 is a diagram showing the BAC contig containing the PG1 gene and the positions of biallelic markers along the contig.

Figure 2 is a graph showing the results of the first screening of a prostate cancer association study and the significance of various biallelic markers as measured by their chi squared and p-values for a low density set of markers.

Figure 3 is a graph showing the results of the first screening of a prostate cancer association study and the significance of various biallelic markers as measured by their chi squared and p-values for a higher density set of markers.

Figure 4 is a table demonstrating the results of an haplotype analysis. Among all the theoretical potential different haplotypes based on 2 to 9 markers, 11 haplotypes showing a strong association with prostate cancer were selected, and their haplotype analysis results are shown here.

Figure 5 is a bar graph demonstrating the results of an experiment evaluating the significance (p-values) of the haplotype analysis shown in Figure 4.

Figure 6A is a table listing the biallelic markers used in the haplotype analysis of Figure 4. Figure 6B is a table listing additional biallelic markers in linkage disequilibrium with the PG1 gene.

Figure 7 is a table listing the positions of exons, splice sites, a stop codon, and a poly A site in the PG1 gene.

Figure 8A is a diagram showing the genomic structure of PG1 in comparison with its most abundant mRNA transcript. Figure 8B is a more detailed diagram showing the genomic structure of PG1, including exons and introns.

Figure 9 is a table listing some of the homologies between the PG1 protein and known proteins.

Figure 10 is a half-tone reproduction of a fluorescence micrograph of the perinuclear/nuclear

expression of PG1 in tumoral (PC3) and normal prostatic cell lines (PNT2). Vector "PG1": includes all the coding exons from exon 1 to 8. For PC3 (upper panel) and PNT2 (lower panel), the nucleus was labelled with Propidium iodide (IP, left panel). Note that EGFP fluorescence was detected in and around the nucleus (GFP, middle panel), as shown when the two pictures were overlapped (right panel).

Figure 11 is a half-tone reproduction of a fluorescence micrograph of the perinuclear/nuclear expression of PG1/1-4 in tumoral (PC3) and normal prostatic cell lines (PNT2). Vector "PG1/1-4" corresponds to an alternative messenger which is due to an alternative splicing, joining exon 1 to exon 4, and resulting in the absence of exons 2 and 3. For PC3 (upper panel) and PNT2 (lower panel), the nucleus was labelled with Propidium iodide (IP, left panel). Note that EGFP fluorescence was detected in and around the nucleus (GFP, middle panel), as shown when the two pictures were overlapped (right panel).

Figure 12 is a half-tone reproduction of a fluorescence micrograph of the perinuclear/nuclear expression of PG1/1-5 in tumoral prostatic cell line (PC3) and cytoplasmic expression of PG1/1-5 in normal prostatic cell line (PNT2). Vector "PG1/1-5" corresponds to an alternative messenger which is due to an alternative splicing, joining exon 1 to exon 5, and resulting in the absence of exons 2, 3 and 4. For PC3 (upper panel) and PNT2 (lower panels), the nucleus was labelled with Propidium iodide (IP). Note that in PC3 cells, EGFP fluorescence was detected in and around the nucleus (GFP, upper middle panel), as shown when the two picture were overlapped (upper right panel). In PNT2A cells, EGFP fluorescence was detected in the cytoplasm (GFP, lower left panel), as shown when the two pictures were overlapped (lower right panel).

Figure 13 is a half-tone reproduction of a fluorescence micrograph of the perinuclear/nuclear expression of a mutated form PG1 (PG1mut229) in normal prostatic cell line (PNT2). Vector "PG1/1-7" includes exons 1 to 6, and corresponds to the mutated form identified in genomic DNA of the prostatic tumoural cell line LNCaP. The nucleus was labelled with Propidium iodide (IP, left panel). EGFP fluorescence was detected in the cytoplasm (GFP, middle panel), as shown when the two pictures were overlapped (lower right panel).

Figure 14 is a diagram of the structure of the 14 alternative splice species found for human PG1 by the exons present. An * indicates that there is a stop codon in frame at that location. An arrow to the right at the right-hand side of a splice species indicates that the open-reading frame continues off of the chart. a space between exons indicates that the exon(s) is missing from that particular alternative splice species. An up arrow indicates that either exon 1bis, 3bis, or 5bis has been inserted depending upon which is indicated. A bracket notation in exon 6, over an exon 6bis notation indicates that the first 60 bases is missing from exon 6, and exon 6bis is therefor present as a truncated form of exon 6.

Figure 15 is a table listing the results of a series of RT-PCR experiments that were performed

on RNA of normal prostate, normal prostatic cell lines (PNT1A, PNT1B and PNT2), and tumoral prostatic cell lines (LnCaPFCG, LnCaPIMB, CaHPV, Du145, PC3, and prostate tumors (ECP5 to ECP24) using all the possible combinations of primers (SEQ ID NOs: 137-178) specific to all of the possible splice junctions or exon borders in human PG1. An NT indicates that the experiment was not performed. An [+] indicates the use of an alternative splice species with exons 1, 3, 4, 7, and 8.

Figure 16 is a graph showing the results of association studies using markers spanning the 650 kb region of the 8p23 locus around PG1, using both single point analysis and haplotyping studies.

Figure 17 is a graph showing an enlarged view of the single point association results within a 160 kb region comprising the PG1 gene.

Figure 18A is a graph showing an enlarged view of the single point association results of 40 kb within the PG1 gene. Figure 18B is a table listing the location of markers within PG1 gene, the two possible alleles at each site. For each marker, the disease-associated allele is indicated first; its frequencies in cases and controls as well as the difference between both are shown; the odd-ratio and the p-value of each individual marker association are also shown.

Figure 19A is a table showing the results of a haplotype analysis study using 4 markers (marker Nos. 4-14, 99-217, 4-66 and 99-221) within the 160 kb region shown in Figure 17. Figure 19B is a table showing the segmented haplotyping results according to the subject's age, and whether the prostate cancer cases were sporadic or familial, using the same markers 4 markers and the same individuals as were used to generate the results in Figure 19A.

Figure 20 is a table listing the haplotyping results and odd ratios for combinations of the 7 markers (99-622 ; 4-77 ; 4-71 ; 4-73 ; 99-598 ; 99-576 ; 4-66) within PG1 gene that were shown in Figure 18 to have p-values more significant than 1.10^{-2} . All of the 2-, 3-, 4-, 5-, 6- and 7-marker haplotypes were tested.

Figure 21 is a graph showing the distribution of statistical significance, as measured by Chi-square values, for each series of possible x-marker haplotypes, (x =2, 3 or 4) using all of the 19 markers listed in Figure 18B.

Detailed Description of the Preferred Embodiment

The practice of the present invention encompasses conventional techniques of chemistry, immunology, molecular biology, biochemistry, protein chemistry, and recombinant DNA technology, which are within the skill of the art. Such techniques are explained fully in the literature. See, e.g., Oligonucleotide Synthesis (M. Gait ed. 1984); Nucleic Acid Hybridization (B. Hames & S. Higgins, eds., 1984); Sambrook, Fritsch & Maniatis, Molecular Cloning: A Laboratory Manual, Second Edition (1989); PCR Technology (H.A. Erlich ed., Stockton Press); R. Scope, Protein Purification Principles and Practice (Springer-Verlag); and the series Methods in Enzymology (S. Colowick and N. Kaplan eds., Academic Press, Inc.).

Definitions

As used interchangeably herein, the terms "nucleic acid" "oligonucleotide", and "polynucleotides" include RNA, DNA, or RNA/DNA hybrid sequences of more than one nucleotide in either single chain or duplex form. The term "nucleotide" as used herein as an adjective to describe molecules comprising RNA, DNA, or RNA/DNA hybrid sequences of any length in single-stranded or duplex form. The term "nucleotide" is also used herein as a noun to refer to individual nucleotides or varieties of nucleotides, meaning a molecule, or individual unit in a larger nucleic acid molecule, comprising a purine or pyrimidine, a ribose or deoxyribose sugar moiety, and a phosphate group, or phosphodiester linkage in the case of nucleotides within an oligonucleotide or polynucleotide. Although the term "nucleotide" is also used herein to encompass "modified nucleotides" which comprise at least one modifications (a) an alternative linking group, (b) an analogous form of purine, (c) an analogous form of pyrimidine, or (d) an analogous sugar, for examples of analogous linking groups, purine, pyrimidines, and sugars see for example PCT publication No. WO 95/04064. However, the polynucleotides of the invention are preferably comprised of greater than 50% conventional deoxyribose nucleotides, and most preferably greater than 90% conventional deoxyribose nucleotides. The polynucleotide sequences of the invention is prepared by any known method, including synthetic, recombinant, ex vivo generation, or a combination thereof, as well as utilizing any purification methods known in the art.

As used herein, the term "purified" does not require absolute purity; rather, it is intended as a relative definition Purification of starting material or natural material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated.

The term "purified" is used herein to describe a polynucleotide or polynucleotide vector of the invention which has been separated from other compounds including, but not limited to other nucleic acids, carbohydrates, lipids and proteins (such as the enzymes used in the synthesis of the polynucleotide), or the separation of covalently closed polynucleotides from linear polynucleotides. A polynucleotide is substantially pure when at least about 50 %, preferably 60 to 75% of a sample exhibits a single polynucleotide sequence and conformation (linear versus covalently close). A substantially pure polynucleotide typically comprises about 50 %, preferably 60 to 90% weight/weight of a nucleic acid sample, more usually about 95%, and preferably is over about 99% pure. Polynucleotide purity or homogeneity is indicated by a number of means well known in the art, such as agarose or polyacrylamide gel electrophoresis of a sample, followed by visualizing a single polynucleotide band upon staining the gel. For certain purposes higher resolution can be provided by using HPLC or other means well known in the art.

The term "polypeptide" refers to a polymer of amino without regard to the length of the polymer; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide.

This term also does not specify or exclude post-expression modifications of polypeptides, for example, polypeptides which include the covalent attachment of glycosyl groups, acetyl groups, phosphate groups, lipid groups and the like are expressly encompassed by the term polypeptide. Also included within the definition are polypeptides which contain one or more analogs of an amino acid
5 (including, for example, non-naturally occurring amino acids, amino acids which only occur naturally in an unrelated biological system, modified amino acids from mammalian systems etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

As used herein, the term "isolated" requires that the material be removed from its original
10 environment (e.g., the natural environment if it is naturally occurring).

The term "purified" is used herein to describe a polypeptide of the invention which has been separated from other compounds including, but not limited to nucleic acids, lipids, carbohydrates and other proteins. A polypeptide is substantially pure when at least about 50 %, preferably 60 to 75% of a sample exhibits a single polypeptide sequence. A substantially pure polypeptide typically comprises
15 about 50 %, preferably 60 to 90% weight/weight of a protein sample, more usually about 95%, and preferably is over about 99% pure. Polypeptide purity or homogeneity is indicated by a number of means well known in the art, such as agarose or polyacrylamide gel electrophoresis of a sample, followed by visualizing a single polypeptide band upon staining the gel. For certain purposes higher resolution can be provided by using HPLC or other means well known in the art.

As used herein, the term "non-human animal" refers to any non-human vertebrate, birds and more usually mammals, preferably primates, farm animals such as swine, goats, sheep, donkeys, and horses, rabbits or rodents, more preferably rats or mice. As used herein, the term "animal" is used to refer to any vertebrate, preferably a mammal. Both the terms "animal" and "mammal" expressly embrace human subjects unless preceded with the term "non-human".

As used herein, the term "antibody" refers to a polypeptide or group of polypeptides which are comprised of at least one binding domain, where an antibody binding domain is formed from the folding of variable domains of an antibody molecule to form three-dimensional binding spaces with an internal surface shape and charge distribution complementary to the features of an antigenic determinant of an antigen., which allows an immunological reaction with the antigen. Antibodies
30 include recombinant proteins comprising the binding domains, as well as fragments, including Fab, Fab', F(ab)₂, and F(ab')₂ fragments.

As used herein, an "antigenic determinant" is the portion of an antigen molecule, in this case an PG1 polypeptide, that determines the specificity of the antigen-antibody reaction. An "epitope" refers to an antigenic determinant of a polypeptide. An epitope can comprise as few as 3 amino acids
35 in a spatial conformation which is unique to the epitope. Generally an epitope consists of at least 6 such amino acids, and more usually at least 8-10 such amino acids. Methods for determining the

amino acids which make up an epitope include x-ray crystallography, 2-dimensional nuclear magnetic resonance, and epitope mapping e.g. the Pepscan method described by H. Mario Geysen et al. 1984, Proc. Natl. Acad. Sci. U.S.A. 81:3998-4002; PCT Publication No. WO 84/03564; and PCT Publication No. WO 84/03506.

5 The term "DNA construct" and "vector" are used herein to mean a purified or isolated polynucleotide that has been artificially designed and which comprises at least two nucleotide sequences that are not found as contiguous nucleotide sequences in their natural environment.

 The terms "trait" and "phenotype" are used interchangeably herein and refer to any visible, detectable or otherwise measurable property of an organism such as symptoms of, or susceptibility to
10 a disease for example. Typically the terms "trait" or "phenotype" are used herein to refer to symptoms of, or susceptibility to cancer or prostate cancer; or to refer to an individual's response to an anti-cancer agent or an anti-prostate cancer agent; or to refer to symptoms of, or susceptibility to side effects to an anticancer agent or an anti-prostate cancer agent.

 The term "allele" is used herein to refer to variants of a nucleotide sequence. A biallelic
15 polymorphism has two forms. Typically the first identified allele is designated as the original allele whereas other alleles are designated as alternative alleles. Diploid organisms is homozygous or heterozygous for an allelic form.

 The term "heterozygosity rate" is used herein to refer to the incidence of individuals in a population, which are heterozygous at a particular allele. In a biallelic system the heterozygosity rate
20 is on average equal to $2P_1(1-P_1)$, where P_1 is the frequency of the least common allele. In order to be useful in genetic studies a genetic marker should have an adequate level of heterozygosity to allow a reasonable probability that a randomly selected person will be heterozygous.

 The term "genotype" as used herein refers the identity of the alleles present in an individual or a sample. In the context of the present invention a genotype preferably refers to the description of the
25 biallelic marker alleles present in an individual or a sample. The term "genotyping" a sample or an individual for a biallelic marker consists of determining the specific allele or the specific nucleotide carried by an individual at a biallelic marker.

 The term "mutation" as used herein refers to a difference in DNA sequence between or among different genomes or individuals which has a frequency below 1%.

30 The term "haplotype" refers to a combination of alleles present in an individual or a sample. In the context of the present invention a haplotype preferably refers to a combination of biallelic marker alleles found in a given individual and which is associated with a phenotype.

 The term "polymorphism" as used herein refers to the occurrence of two or more alternative genomic sequences or alleles between or among different genomes or individuals. "Polymorphic"
35 refers to the condition in which two or more variants of a specific genomic sequence can be found in a population. A "polymorphic site" is the locus at which the variation occurs. A single nucleotide

polymorphism is a single base pair change. Typically a single nucleotide polymorphism is the replacement of one nucleotide by another nucleotide at the polymorphic site. Deletion of a single nucleotide or insertion of a single nucleotide, also give rise to single nucleotide polymorphisms. In the context of the present invention "single nucleotide polymorphism" preferably refers to a single
5 nucleotide substitution. Typically, between different genomes or between different individuals, the polymorphic site is occupied by two different nucleotides.

The terms "biallelic polymorphism" and "biallelic marker" are used interchangeably herein to refer to a nucleotide polymorphism having two alleles at a fairly high frequency in the population. A "biallelic marker allele" refers to the nucleotide variants present at a biallelic marker site. Usually a
10 biallelic marker is a single nucleotide polymorphism. However, less commonly there are also insertions and deletions of up to 5 nucleotides which constitute biallelic markers for the purposes of the present invention. Typically the frequency of the less common allele of the biallelic markers of the present invention has been validated to be greater than 1%, preferably the frequency is greater than 10%, more preferably the frequency is at least 20% (i.e. heterozygosity rate of at least 0.32),
15 even more preferably the frequency is at least 30% (i.e. heterozygosity rate of at least 0.42). A biallelic marker wherein the frequency of the less common allele is 30% or more is termed a "high quality biallelic marker."

The location of nucleotides in a polynucleotide with respect to the center of the polynucleotide are described herein in the following manner. When a polynucleotide has an odd
20 number of nucleotides, the nucleotide at an equal distance from the 3' and 5' ends of the polynucleotide is considered to be "at the center" of the polynucleotide, and any nucleotide immediately adjacent to the nucleotide at the center, or the nucleotide at the center itself is considered to be "within 1 nucleotide of the center." With an odd number of nucleotides in a polynucleotide any of the five nucleotides positions in the middle of the polynucleotide would be considered to be within
25 2 nucleotides of the center, and so on. When a polynucleotide has an even number of nucleotides, there would be a bond and not a nucleotide at the center of the polynucleotide. Thus, either of the two central nucleotides would be considered to be "within 1 nucleotide of the center" and any of the four nucleotides in the middle of the polynucleotide would be considered to be "within 2 nucleotides of the center", and so on.

30 The term "upstream" is used herein to refer to a location which is toward the 5' end of the polynucleotide from a specific reference point.

The terms "base paired" and "Watson & Crick base paired" are used interchangeably herein to refer to nucleotides which can be hydrogen bonded to one another by virtue of their sequence identities in a manner like that found in double-helical DNA with thymine or uracil residues linked to
35 adenine residues by two hydrogen bonds and cytosine and guanine residues linked by three hydrogen bonds (See Stryer, L., *Biochemistry*, 4th edition, 1995).

The terms "complementary" or "complement thereof" are used herein to refer to the sequences of polynucleotides which is capable of forming Watson & Crick base pairing with another specified polynucleotide throughout the entirety of the complementary region. This term is applied to pairs of polynucleotides based solely upon their sequences and not any particular set of conditions under which the two polynucleotides would actually bind.

As used herein the term "PG1-related biallelic marker" relates to a set of biallelic markers in linkage disequilibrium with PG1. The term PG1-related biallelic marker includes all of the biallelic markers used in the initial association studies shown below in Section I.D., including those biallelic markers contained in SEQ ID NOs: 21-38 and 57-62. The term PG1-related biallelic marker encompasses all of the following polymorphisms positioned in SEQ ID 179, and listed by internal reference number, including:

- 5-63-169 G or C in position 2159;
- 5-63-453 C or T in position 2443; 99-622-95 T or C in position 4452;
- 99-621-215 T or C in position 5733; 99-619-141 G or A in position 8438;
- 4-76-222 deletion of GT in position 11843; 4-76-361 C or T in position 11983;
- 4-77-151 G or C in position 12080; 4-77-294 A or G in position 12221;
- 4-71-33 G or T in position 12947; 4-71-233 A or G in position 13147;
- 4-71-280 G or A in position 13194; 4-71-396 G or C in position 13310;
- 4-72-127 A or G in position 13342; 4-72-152 A or G in position 13367;
- 4-72-380 deletion of A in position 13594; 4-73-134 G or C in position 13680;
- 4-73-356 G or C in position 13902; 99-610-250 T or C in position 16231;
- 99-610-93 A or T in position 16388; 99-609-225 A or T in position 17608;
- 4-90-27 A or C in position 18034; 4-90-283 A or C in position 18290;
- 99-607-397 T or C in position 18786; 99-602-295 deletion of A in position 22835;
- 99-602-258 T or C in position 22872;
- 99-600-492 deletion of TATTG in position 25183;
- 99-600-483 T or G in position 25192; 5-23-288 A or G in position 25614;
- 99-598-130 T or C in position 26911; 99-592-139 A or T in position 32703;
- 99-217-277 C or T in position 34491; 5-47-284 A or G in position 34756;
- 99-589-267 T or G in position 34934; 99-589-41 G or C in position 35160;
- 99-12899-307 C or T in position 39897; 4-12-68 A or G in position 40598;
- 99-582-263 T or C in position 40816; 99-582-132 T or C in position 40947;
- 99-576-421 G or C in position 45783; 4-13-51 C or T in position 47929;
- 4-13-328 A or T in position 48206; 4-13-329 G or C in position 48207;
- 99-12903-381 C or T in position 49282; 5-56-208 A or G in position 50037;
- 5-56-225 A or G in position 50054; 5-56-272 A or G in position 50101;
- 5-56-391 G or T in position 50220; 4-61-269 A or G in position 50440;

- 4-61-391 A or G in position 50562; 4-63-99 A or G in position 50653;
 4-62-120 A or G in position 50660; 4-62-205 A or G in position 50745;
 4-64-113 A or T in position 50885; 4-65-104 A or G in position 51249;
 5-28-300 A or G in position 51333; 5-50-269 C or T in position 51435;
 5 4-65-324 C or T in position 51468; 5-71-129 G or C in position 51515;
 5-50-391 G or C in position 51557; 5-71-180 A or G in position 51566;
 4-67-40 C or T in position 51632; 5-71-280 A or C in position 51666;
 5-58-167 A or G in position 52016; 5-30-325 C or T in position 52096;
 5-58-302 A or T in position 52151; 5-31-178 A or G in position 52282;
 10 5-31-244 A or G in position 52348; 5-31-306 deletion of A in position 52410;
 5-32-190 C or T in position 52524; 5-32-246 C or T in position 52580;
 5-32-378 deletion of A in position 52712; 5-53-266 G or C in position 52772;
 5-60-158 C or T in position 52860; 5-60-390 A or G in position 53092;
 5-68-272 G or C in position 53272; 5-68-385 A or T in position 53389;
 15 5-66-53 deletion of GA in position 53511; 5-66-142 G or C in position 53600;
 5-66-207 A or G in position 53665; 5-37-294 A or G in position 53815;
 5-62-163 insertion of A in position 54365; 5-62-340 A or T in position 54541; and the compliments
 thereof. The term PG1-related biallelic marker also includes all of the following biallelic markers
 listed by internal reference number, and two SEQ ID NOs each of which contains a 47-mers with one
 20 of the two alternative bases at position 24:
 4-14-107 of SEQ ID NOs 185 and 262; 4-14-317 of SEQ ID NOs 186 and 263; 4-14-35 of
 SEQ ID NOs 187 and 264; 4-20-149 of SEQ ID NOs 188 and 265;
 4-20-77 of SEQ ID NOs 189 and 266; 4-22-174 of SEQ ID NOs 190 and 267;
 4-22-176 of SEQ ID NOs 191 and 268; 4-26-60 of SEQ ID NOs 192 and 269;
 25 4-26-72 of SEQ ID NOs 193 and 270; 4-3-130 of SEQ ID NOs 194 and 271;
 4-38-63 of SEQ ID NOs 195 and 272;
 4-38-83 of SEQ ID NOs 196 and 273; 4-4-152 of SEQ ID NOs 197 and 274;
 4-4-187 of SEQ ID NOs 198 and 275; 4-4-288 of SEQ ID NOs 199 and 276;
 4-42-304 of SEQ ID NOs 200 and 277; 4-42-401 of SEQ ID NOs 201 and 278;
 30 4-43-328 of SEQ ID NOs 202 and 279; 4-43-70 of SEQ ID NOs 203 and 280;
 4-50-209 of SEQ ID NOs 204 and 281; 4-50-293 of SEQ ID NOs 205 and 282;
 4-50-323 of SEQ ID NOs 206 and 283; 4-50-329 of SEQ ID NOs 207 and 284;
 4-50-330 of SEQ ID NOs 208 and 285; 4-52-163 of SEQ ID NOs 209 and 286;
 4-52-88 of SEQ ID NOs 210 and 287; 4-53-258 of SEQ ID NOs 211 and 288;
 35 4-54-283 of SEQ ID NOs 212 and 289; 4-54-388 of SEQ ID NOs 213 and 290;
 4-55-70 of SEQ ID NOs 214 and 291; 4-55-95 of SEQ ID NOs 215 and 292;

- 4-56-159 of SEQ ID NOs 216 and 293; 4-56-213 of SEQ ID NOs 217 and 294;
 4-58-289 of SEQ ID NOs 218 and 295; 4-58-318 of SEQ ID NOs 219 and 296;
 4-60-266 of SEQ ID NOs 220 and 297; 4-60-293 of SEQ ID NOs 221 and 298;
 4-84-241 of SEQ ID NOs 222 and 299; 4-84-262 of SEQ ID NOs 223 and 300;
 5 4-86-206 of SEQ ID NOs 224 and 301; 4-86-309 of SEQ ID NOs 225 and 302;
 4-88-349 of SEQ ID NOs 226 and 303; 4-89-87 of SEQ ID NOs 227 and 304;
 99-123-184 of SEQ ID NOs 228 and 305; 99-128-202 of SEQ ID NOs 229 and 306;
 99-128-275 of SEQ ID NOs 230 and 307; 99-128-313 of SEQ ID NOs 231 and 308; 99-128-60 of
 SEQ ID NOs 232 and 309; 99-12907-295 of SEQ ID NOs 233 and 310; 99-130-58 of SEQ ID NOs
 10 234 and 311; 99-134-362 of SEQ ID NOs 235 and 312;
 99-140-130 of SEQ ID NOs 236 and 313; 99-1462-238 of SEQ ID NOs 237 and 314; 99-147-181 of
 SEQ ID NOs 238 and 315; 99-1474-156 of SEQ ID NOs 239 and 316; 99-1474-359 of SEQ ID
 NOs 240 and 317; 99-1479-158 of SEQ ID NOs 241 and 318; 99-1479-379 of SEQ ID NOs 242 and
 319; 99-148-129 of SEQ ID NOs 243 and 320; 99-148-132 of SEQ ID NOs 244 and 321; 99-148-139
 15 of SEQ ID NOs 245 and 322;
 99-148-140 of SEQ ID NOs 246 and 323; 99-148-182 of SEQ ID NOs 247 and 324;
 99-148-366 of SEQ ID NOs 248 and 325; 99-148-76 of SEQ ID NOs 249 and 326;
 99-1480-290 of SEQ ID NOs 250 and 327; 99-1481-285 of SEQ ID NOs 251 and 328; 99-1484-101 of
 SEQ ID NOs 252 and 329; 99-1484-328 of SEQ ID NOs 253 and 330; 99-1485-251 of SEQ ID NOs
 20 254 and 331; 99-1490-381 of SEQ ID NOs 255 and 332; 99-1493-280 of SEQ ID NOs 256 and
 333; 99-151-94 of SEQ ID NOs 257 and 334;
 99-211-291 of SEQ ID NOs 258 and 335; 99-213-37 of SEQ ID NOs 259 and 336;
 99-221-442 of SEQ ID NOs 260 and 337; 99-222-109 of SEQ ID NOs 261 and 338; and the
 compliments thereof.
- 25 The term "non-genic" is used herein to describe PG1-related biallelic markers, as well as
 polynucleotides and primers which do not occur in the human PG1 genomic sequence of SEQ ID NO:
 179. The term "genic" is used herein to describe PG1-related biallelic markers as well as
 polynucleotides and primers which do occur in the human PG1 genomic sequence of SEQ ID NO:
 179.
- 30 The terms "an anti-cancer agent" refers to a drug or a compound that is capable of reducing
 the growth rate, rate of metastasis, or viability of tumor cells in a mammal, is capable of reducing the
 size or eliminating tumors in a mammal, or is capable of increasing the average life span of a mammal
 or human with cancer. Anti-cancer agents also include compounds which are able to reduce the risk
 of cancer developing in a population, particularly a high risk population. The terms "an anti-prostate
 35 cancer agent" is an anti-cancer agent that has these effects on cells or tumors that are derived from
 prostate cancer cells.

The terms "response to an anti-cancer agent" and "response to an anti-prostate cancer agent" refer to drug efficacy, including but not limited to ability to metabolize a compound, to the ability to convert a pro-drug to an active drug, and to the pharmacokinetics (absorption, distribution, elimination) and the pharmacodynamics (receptor-related) of a drug in an individual.

5 The terms "side effects to an anti-cancer agent" and "side effects to an anti-prostate cancer agent" refer to adverse effects of therapy resulting from extensions of the principal pharmacological action of the drug or to idiosyncratic adverse reactions resulting from an interaction of the drug with unique host factors. These side effects include, but are not limited to, adverse reactions such as dermatological, hematological or hepatological toxicities and further includes gastric and intestinal
10 ulceration, disturbance in platelet function, renal injury, nephritis, vasomotor rhinitis with profuse watery secretions, angioneurotic edema, generalized urticaria, and bronchial asthma to laryngeal edema and bronchoconstriction, hypotension, sexual dysfunction, and shock.

As used herein the term "homology" refers to comparisons between protein and/or nucleic acid sequences and is evaluated using any of the variety of sequence comparison algorithms and
15 programs known in the art. Such algorithms and programs include, but are by no means limited to, TBLASTN, BLASTP, FASTA, TFASTA, and CLUSTALW (Pearson and Lipman, 1988, Proc. Natl. Acad. Sci. USA 85(8):2444-2448; Altschul et al., 1990, J. Mol. Biol. 215(3):403-410; Thompson et al., 1994, Nucleic Acids Res. 22(2):4673-4680; Higgins et al., 1996, Methods Enzymol. 266:383-402; Altschul et al., 1990, J. Mol. Biol. 215(3):403-410; Altschul et al., 1993, Nature Genetics 3:266-272).
20 In a particularly preferred embodiment, protein and nucleic acid sequence homologies are evaluated using the Basic Local Alignment Search Tool ("BLAST") which is well known in the art (see, e.g., Karlin and Altschul, 1990, Proc. Natl. Acad. Sci. USA 87:2267-2268; Altschul et al., 1990, J. Mol. Biol. 215:403-410; Altschul et al., 1993, Nature Genetics 3:266-272; Altschul et al., 1997, Nuc. Acids Res. 25:3389-3402). In particular, five specific BLAST programs are used to perform the following
25 task:

- (1) BLASTP and BLAST3 compare an amino acid query sequence against a protein sequence database;
- (2) BLASTN compares a nucleotide query sequence against a nucleotide
30 sequence database;
- (3) BLASTX compares the six-frame conceptual translation products of a query nucleotide sequence (both strands) against a protein sequence database;
- (4) TBLASTN compares a query protein sequence against a nucleotide sequence database translated in all six reading frames (both strands); and

- (5) TBLASTX compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

The BLAST programs identify homologous sequences by identifying similar segments, which are referred to herein as "high-scoring segment pairs," between a query amino or nucleic acid sequence and a test sequence which is preferably obtained from a protein or nucleic acid sequence database. High-scoring segment pairs are preferably identified (i.e., aligned) by means of a scoring matrix, many of which are known in the art. Preferably, the scoring matrix used is the BLOSUM62 matrix (Gonnet et al., 1992, Science 256:1443-1445; Henikoff and Henikoff, 1993, Proteins 17:49-61). Less preferably, the PAM or PAM250 matrices may also be used (see, e.g., Schwartz and Dayhoff, eds., 1978, Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure, Washington: National Biomedical Research Foundation). The BLAST programs evaluate the statistical significance of all high-scoring segment pairs identified, and preferably selects those segments which satisfy a user-specified threshold of significance, such as a user-specified percent homology. Preferably, the statistical significance of a high-scoring segment pair is evaluated using the statistical significance formula of Karlin (see, e.g., Karlin and Altschul, 1990, Proc. Natl. Acad. Sci. USA 87:2267-2268).

I. ISOLATION AND CHARACTERIZATION OF THE PG1 GENE AND PROTEINS

I.A. The 8p23 Region- LOH Studies: Implications of 8p23 Region in Distinct Cancer Types

Substantial amounts of LOH data support the hypothesis that genes associated with distinct cancer types are located within 8p23 region of the human genome. Emi et al., demonstrated the implication of 8p23.1-8p21.3 region in cases of hepatocellular carcinoma, colorectal cancer, and non-small cell lung cancer. (Emi M, Fujiwara Y, Nakajima T, Tsuchiya E, Tsuda H, Hirohashi S, Maeda Y, Tsuruta K, Miyaki M, Nakamura Y, Cancer Res. 1992 Oct 1; 52(19): 5368-5372) Yaremko, et al., showed the existence of two major regions of LOH for chromosome 8 markers in a sample of 87 colorectal carcinomas. The most prominent loss was found for 8p23.1-pter, where 45% of informative cases demonstrated loss of alleles. (Yaremko ML, Wasylyshyn ML, Paulus KL, Michelassi F, Westbrook CA, Genes Chromosomes Cancer 1994 May;10(1):1-6). Scholnick et al. demonstrated the existence of three distinct regions of LOH for the markers of chromosome 8 in cases of squamous cell carcinoma of the supraglottic larynx. They showed that the allelic loss of 8p23 marker D8S264 serves as a statistically significant, independent predictor of poor prognosis for patients with supraglottic squamous cell carcinoma. (Scholnick SB, Haughey BH, Sunwoo JB, el-Mofty SK, Baty JD, Piccirillo JF, Zequeira MR, J. Natl. Cancer Inst. 1996 Nov 20; 88(22): 1676-1682 and Sunwoo JB, Holt MS, Radford DM, Deeker C, Scholnick S, Genes Chromosomes Cancer 1996 Jul; 16(3):164-169).

In other studies, Nagai et al. demonstrated the highest loss of heterozygosity in the specific region of 8p23 by genome wide scanning of LOH in 120 cases of hepatocellular carcinoma (HCC). (Nagai H, Pineau P, Tiollais P, Buendia MA, Dejean A, *Oncogene* 1997 Jun 19; 14(24): 2927-2933). Gronwald et al. demonstrated 8p23-pter loss in renal clear cell carcinomas. (Gronwald J, Storkel S, Holtgreve-Grez H, Hadaczek P, Brinkschmidt C, Jauch A, Lubinski J, Cremer, *Cancer Res.* 1997 Feb 1; 57(3): 481-487).

The same region is involved in specific cases of prostate cancer. Matsuyama et al. showed the specific deletion of the 8p23 band in prostate cancer cases, as monitored by FISH with D8S7 probe. (Matsuyama H, Pan Y, Skoog L, Tribukait B, Naito K, Ekman P, Lichter P, Bergerheim US *Oncogene* 1994 Oct; 9(10): 3071-3076). They were able to document a substantial number of cases with deletions of 8p23 but retention of the 8p22 marker LPL. Moreover, Ichikawa et al. deduced the existence of a prostate cancer metastasis suppressor gene and localized it to 8p23-q12 by studies of metastasis suppression in highly metastatic rat prostate cells after transfer of human chromosomes. (Ichikawa T, Nihei N, Kuramochi H, Kawana Y, Killary AM, Rinker-Schaeffer CW, Barrett JC, Isaacs JT, Kugoh H, Oshimura M, Shimazaki J, *Prostate Suppl.* 1996; 6: 31-35).

Recently Washburn et al. were able to find substantial numbers of tumors with the allelic loss specific to 8p23 by LOH studies of 31 cases of human prostate cancer. (Washburn J, Woino K, and Macoska J, *Proceedings of American Association for Cancer Research*, March 1997; 38). In these samples they were able to define the minimal overlapping region with deletions covering genetic interval D8S262-D8S277.

Linkage Analysis Studies: Search for Prostate Cancer

Linked Regions on Chromosome 8

Microsatellite markers mapping to chromosome 8 were used by the inventors to perform linkage analysis studies on 194 individuals issued from 47 families affected with prostate cancer. While multiple point analysis led to weak linkage results, two point lod score analysis led to non significant results, as shown below.

Two point lod (parametric analysis)

MARKER	Distance (cM)	Z(lod) scores
D8S1742	0.8	-0.13
D8S561		-0.07

# of families analyzed	47
Total # of individuals genotyped	194
Total # of affected individuals genotyped	122

In view of the non-significant results obtained with linkage analysis, a new mapping approach based on linkage disequilibrium of biallelic markers was utilised to identify genes responsible for sporadic cases of prostate cancer.

I.B. Linkage Disequilibrium Using Biallelic Markers To Identify Candidate Loci Responsible

5 For Disease

Linkage Disequilibrium

Once a chromosomal region has been identified as potentially harboring a candidate gene associated with a sporadic trait, an excellent approach to refine the candidate gene's location within the identified region is to look for statistical associations between the trait and some marker genotype
10 when comparing an affected (trait⁺) and a control (trait⁻) population.

Association studies have most usually relied on the use of biallelic markers. Biallelic markers are genome-derived polynucleotides that exhibit biallelic polymorphism at one single base position. By definition, the lowest allele frequency of a biallelic polymorphism is 1%; sequence variants that show allele frequencies below 1% are called rare mutations. There are potentially more than 10⁷
15 biallelic markers lying along the human genome.

Association studies seek to establish correlations between traits and genetic markers and are based on the phenomenon of linkage disequilibrium (LD). LD is defined as the trend for alleles at nearby loci on haploid genomes to correlate in the population. If two genetic loci lie on the same chromosome, then sets of alleles on the same chromosomal segment (i.e., haplotypes) tend to be
20 transmitted as a block from generation to generation. When not broken up by recombination, haplotypes can be tracked not only through pedigrees but also through populations. The resulting phenomenon at the population level is that the occurrence of pairs of specific alleles at different loci on the same chromosome is not random, and the deviation from random is called linkage disequilibrium.

25 Since results generated by association studies are essentially based on the quantitative calculation of allele frequencies, they best apply to the analysis of germline mutations. This is mainly due to the fact that allelic frequencies are difficult to quantify within tumor tissue samples because of the usual presence of normal cells within the studied tumor samples. Association studies applied to cancer genetics will therefore be best suited to the identification of tumor suppressor genes.

30 Trait Localization by Linkage Disequilibrium Mapping

Any gene responsible or partly responsible for a given trait will be in LD with some flanking markers. To map such a gene, specific alleles of these flanking markers which are associated with the gene or genes responsible for the trait are identified. Although the following discussion of techniques for finding the gene or genes associated with a particular trait using linkage disequilibrium mapping,
35 refers to locating a single gene which is responsible for the trait, it will be appreciated that the same techniques may also be used to identify genes which are partially responsible for the trait.

Association studies is conducted within the general population (as opposed to the linkage analysis techniques discussed above which are limited to studies performed on related individuals in one or several affected families).

Association between a biallelic marker A and a trait T may primarily occur as a result of three possible relationships between the biallelic marker and the trait. First, allele *a* of biallelic marker A is directly responsible for trait T (e.g., Apo E e4 allele and Alzheimer's disease). However, since the majority of the biallelic markers used in genetic mapping studies are selected randomly, they mainly map outside of genes. Thus, the likelihood of allele *a* being a functional mutation directly related to trait T is therefore very low.

An association between a biallelic marker A and a trait T may also occur when the biallelic marker is very closely linked to the trait locus. In other words, an association occurs when allele *a* is in linkage disequilibrium with the trait-causing allele. When the biallelic marker is in close proximity to a gene responsible for the trait, more extensive genetic mapping will ultimately allow a gene to be discovered near the marker locus which carries mutations in people with trait T (i.e. the gene responsible for the trait or one of the genes responsible for the trait). As will be further exemplified below using a group of biallelic markers which are in close proximity to the gene responsible for the trait, the location of the causal gene can be deduced from the profile of the association curve between the biallelic markers and the trait. The causal gene will be found in the vicinity of the marker showing the highest association with the trait.

Finally, an association between a biallelic marker and a trait may occur when people with the trait and people without the trait correspond to genetically different subsets of the population who, coincidentally, also differ in the frequency of allele *a* (population stratification). This phenomenon is avoided by using large heterogeneous samples.

Association studies are particularly suited to the efficient identification of susceptibility genes that present common polymorphisms, and are involved in multifactorial traits whose frequency is relatively higher than that of diseases with monofactorial inheritance.

Application of Linkage Disequilibrium Mapping to Candidate Gene Identification

The general strategy of association studies using a set of biallelic markers, is to scan two pools of individuals (affected individuals and unaffected controls) characterized by a well defined phenotype in order to measure the allele frequencies for a number of the chosen markers in each of these pools. If a positive association with a trait is identified using an array of biallelic markers having a high enough density, the causal gene will be physically located in the vicinity of the associated markers, since the markers showing positive association to the trait are in linkage disequilibrium with the trait locus. Regions harboring a gene responsible for a particular trait which

are identified through association studies using high density sets of biallelic markers will, on average, be 20 - 40 times shorter in length than those identified by linkage analysis.

Once a positive association is confirmed as described above, BACs (bacterial artificial chromosomes) obtained from human genomic libraries, constructed as described below, harboring the
5 markers identified in the association analysis are completely sequenced.

Once a candidate region has been sequenced and analyzed, the functional sequences within the candidate region (exons and promoters, and other potential regulatory regions) are scanned for mutations which are responsible for the trait by comparing the sequences of a selected number of controls and affected individuals using appropriate software. Candidate mutations are further
10 confirmed by screening a larger number of affected individuals and controls using the microsequencing techniques described below.

Candidate mutations are identified as follows. A pair of oligonucleotide primers is designed in order to amplify the sequences of every predicted functional region. PCR amplification of each predicted functional sequence is carried out on genomic DNA samples from affected patients and
15 unaffected controls. Amplification products from genomic PCR are subjected to automated dideoxy terminator sequencing reactions and electrophoresed on ABI 377 sequencers. Following gel image analysis and DNA sequence extraction, the sequence data are automatically analyzed to detect the presence of sequence variations among affected cases and unaffected controls. Sequences are systematically verified by comparing the sequences of both DNA strands of each individual.

20 Polymorphisms are then verified by screening a larger population of affected individuals and controls using the microsequencing technique described below in an individual test format. Polymorphisms are considered as candidate mutations when present in affected individuals and controls at frequencies compatible with the expected association results.

Association Studies: Statistical Analysis and Haplotyping

25 As mentioned above, linkage analysis typically localizes a disease gene to a chromosomal region of several megabases. Further refinement in location requires the analysis of additional families in order to increase the number of recombinants. However, this approach becomes unfeasible because recombination is rarely observed even within large pedigrees (Boehnke, M, 1994, Am. J. Hum. Genet. 55: 379-390).

30 Linkage disequilibrium, the nonrandom association of alleles at linked loci, may offer an alternative method of obtaining additional recombinants. When a chromosome carrying a mutant allele of a gene responsible for a given trait is first introduced into a population as a result of either mutation or migration, the mutant allele necessarily resides on a chromosome having a unique set of linked markers (haplotype). Consequently, there is complete disequilibrium between these markers
35 and the disease mutation: the disease mutation is present only linked to a specific set of marker alleles. Through subsequent generations, recombinations occur between the disease mutation and

these marker polymorphisms, resulting in a gradual disappearance of disequilibrium. The degree of disequilibrium dissipation depends on the recombination frequency, so the markers closest to the disease gene will tend to show higher levels of disequilibrium than those that are farther away (Jorde LB, 1995, Am. J. Hum. Genet. 56: 11-14). Because linkage disequilibrium patterns in a present-day
5 population reflect the action of recombination through many past generations, disequilibrium analysis effectively increases the sample of recombinants. Thus the mapping resolution achieved through the analysis of linkage disequilibrium patterns is much higher than that of linkage analysis.

In practice, in order to define the regions bearing a candidate gene, the affected and control populations are genotyped using an appropriate number of biallelic markers (at a density of 1 marker
10 every 50-150 kilobases). Then, a marker/trait association study is performed that compares the genotype frequency of each biallelic marker in the affected and control populations by means of a chi square statistical test (one degree of freedom).

After the first screening, additional markers within the region showing positive association are genotyped in the affected and control populations. Two types of complementary analysis are then
15 performed. First, a marker/trait association study (as described above) is performed to refine the location of the gene responsible for the trait. In addition, a haplotype association analysis is performed to define the frequency and the type of the ancestral/preferential carrier haplotype. Haplotype analysis, by combining the informativeness of a set of biallelic markers increases the power of the association analysis, allowing false positive and/or negative data that may result from the
20 single marker studies to be eliminated.

The haplotype analysis is performed by estimating the frequencies of all possible haplotypes for a given set of biallelic markers in the case and control populations, and comparing these frequencies by means of a chi square statistical test (one degree of freedom). Haplotype estimations are performed by applying the Expectation-Maximization (EM) algorithm (Excoffier L & Slatkin M,
25 1995, Mol. Biol. Evol. 12: 921-927), using the EM-HAPLO program (Hawley ME, Pakstis AJ & Kidd KK, 1994, Am. J. Phys. Anthropol. 18: 104). The EM algorithm is used to estimate haplotype frequencies in the case when only genotype data from unrelated individuals are available. The EM algorithm is a generalized iterative maximum likelihood approach to estimation that is useful when data are ambiguous and/or incomplete.

30 The application of biallelic marker based linkage disequilibrium analysis to the 8p23 region to identify a gene associated with prostate cancer is described below.

I.C. Application of Linkage Disequilibrium Mapping to the 8p23 Region

YAC Contig Construction in 8p23 Region

First, a YAC contig which contains the 8p23 region was constructed as follows. The CEPH-
35 Genethon YAC map for the entire human genome (Chumakov I.M. et al. A YAC contig map of the human genome, Nature, 377 Supp.: 175-297, 1995) was used for detailed contig building in the region

around D8S262 and D8S277 genetic markers. Screening data available for regional genetic markers D8S1706, D8S277, D8S1742, D8S518, D8S262, D8S1798, D8S1140, D8S561 and D8S1819 were used to select the following set of CEPH YACs, localized within this region: 832_g_12, 787_c_11, 920_h_7, 807_a_1, 842_b_1, 745_a_3, 910_d_3, 879_f_11, 918_c_6, 764_c_7, 910_f_12, 967_c_11, 856_d_8, 792_a_6, 812_h_4, 873_c_8, 930_a_2, 807_a_1, 852_d_10. This set of YACs was tested by PCR with the above mentioned genetic markers as well as with other publicly available markers supposedly located within the 8p23 region. As a result of these studies, a YAC STS contig map was generated around genetic markers D8S262 and D8S277. The two CEPH YACs, 920_h_7 (1170 kb insert size) and 910_f_12 (1480 kb insert size) constitute a minimal tiling path in this region, with an estimated size of ca. 2 Megabases.

During this mapping effort, the following publicly known STS markers were precisely located within the contig: WI-14718, WI-3831, D8S1413E, WI-8327, WI-3823, ND4.

BAC Contig Construction Covering D8S262-D8S277

Fragment Within 8p23 Region of the Human Genome

Following construction of the YAC contig, a BAC contig was constructed as follows. BAC libraries were obtained as described in Woo et al. Nucleic Acids Res., 1994, 22, 4922-4931. Briefly, two different whole human genome libraries were produced by cloning BamHI or HindIII partially digested DNA from a lymphoblastoid cell line (derived from individual N°8445, CEPH families) into the pBeloBAC11 vector (Kim et al. Genomics, 1996, 34, 213-218). The library produced with the BamHI partial digestion contained 110,000 clones with an average insert size of 150 kb, which corresponds to 5 human haploid genome equivalents. The library prepared with the HindIII partial digestion corresponds to 3 human genome equivalents with an average insert size of 150 kb.

BAC Screening

The human genomic BAC libraries obtained as described above were screened with all of the above mentioned STSs. DNA from the clones in both libraries was isolated and pooled in a three dimensional format ready for PCR screening with the above mentioned STSs using high throughput PCR methods (Chumakov et al., Nature 1995, 377: 175-298). Briefly, three dimensional pooling consists in rearranging the samples to be tested in a manner which allows the number of PCR reactions required to screen the clones with STSs to be reduced by at least 100 fold, as compared to screening each clone individually. PCR amplification products were detected by conventional agarose gel electrophoresis combined with automated image capturing and processing.

In a final step, STS-positive clones were checked individually. Subchromosomal localization of BACs was systematically verified by fluorescence in situ hybridization (FISH), performed on metaphasic chromosomes as described by Cherif et al. Proc. Natl. Acad. Sci. USA 1990, 87: 6639-6643.

BAC insert size was determined by Pulsed Field Gel Electrophoresis after digestion with restriction enzyme NotI.

BAC Contig Analysis

The ordered BACs selected by STS screening and verified by FISH, were assembled into
5 contigs and new markers were generated by partial sequencing of insert ends from some of them. These markers were used to fill the gaps in the contig of BAC clones covering the chromosomal region around D8S277, having an estimated size of 2 megabases. Selected BAC clones from the contig were subcloned and sequenced.

BAC Subcloning

10 Each BAC human DNA was first extracted using the alkaline lysis procedure and then sheared by sonication. The obtained DNA fragments were end-repaired and electrophoresed on preparative agarose gels. The fragments in the desired size range were isolated from the gel, purified and ligated to a linearized, dephosphorylated, blunt-ended plasmid cloning vector (pBluescript II Sk (+)). Example 1 describes the BAC subcloning procedure.

Example 1

15 The cells obtained from three liters overnight culture of each BAC clone were treated by alkaline lysis using conventional techniques to obtain the BAC DNA containing the genomic DNA inserts. After centrifugation of the BAC DNA in a cesium chloride gradient, ca. 50µg of BAC DNA was purified. 5-10µg of BAC DNA was sonicated using three distinct conditions, to obtain fragments
20 of the desired size. The fragments were treated in a 50 µl volume with two units of Vent polymerase for 20 min at 70°C, in the presence of the four deoxytriphosphates (100µM). The resulting blunt-ended fragments were separated by electrophoresis on low-melting point 1% agarose gels (60 Volts for 3 hours). The fragments were excised from the gel and treated with agarase. After chloroform extraction and dialysis on Microcon 100 columns, DNA in solution was adjusted to a 100 ng/µl
25 concentration. A ligation was performed overnight by adding 100 ng of BAC fragmented DNA to 20 ng of pBluescript II Sk (+) vector DNA linearized by enzymatic digestion, and treated by alkaline phosphatase. The ligation reaction was performed in a 10 µl final volume in the presence of 40 units/µl T4 DNA ligase (Epicentre). The ligated products were electroporated into the appropriate cells (ElectroMAX E.coli DH10B cells). IPTG and X-gal were added to the cell mixture, which was
30 then spread on the surface of an ampicillin-containing agar plate. After overnight incubation at 37°C, recombinant (white) colonies were randomly picked and arrayed in 96 well microplates for storage and sequencing.

Partial Sequencing of BACs

At least 30 of the obtained BAC clones were sequenced by the end pair-wise method (500 bp
35 sequence from each end) using a dye-primer cycle sequencing procedure. Pair-wise sequencing was performed until a map allowing the relative positioning of selected markers along the corresponding

DNA region was established. Example 2 describes the sequencing and ordering of the BAC inserts.

Example 2

The subclone inserts were amplified by PCR on overnight bacterial cultures, using vector primers flanking the insertions. The insert extremity sequences (on average 500 bases at each end) were determined by fluorescent automated sequencing on ABI 377 sequencers, with a ABI Prism DNA Sequencing Analysis software (2.1.2 version).

The sequence fragments from BAC subclones were assembled using Gap4 software from R. Staden (Bonfield et al. 1995). This software allows the reconstruction of a single sequence from sequence fragments. The sequence deduced from the alignment of different fragments is called the consensus sequence. We used directed sequencing techniques (primer walking) to complete sequences and link contigs.

Figure 1 shows the overlapping BAC subclones (labeled BAC) which make up the assembled contig and the positions of the publicly known STS markers along the contig.

Identification of Biallelic Markers Lying Along the BAC Contig

Following assembly of the BAC contig, biallelic markers lying along the contig were then identified. Given that the assessed distribution of informative biallelic markers in the human genome (biallelic polymorphisms with a heterozygosity rate higher than 42%) is one in 2.5 to 3 kb, six 500 bp genomic fragments have to be screened in order to identify 1 biallelic marker. Six pairs of primers per potential marker, each one defining a ca. 500 bp amplification fragment, were derived from the above mentioned BAC partial sequences. All primers contained a common upstream oligonucleotide tail enabling the easy systematic sequencing of the resulting amplification fragments. Amplification of each BAC-derived sequence was carried out on pools of DNA from ca. 100 individuals. The conditions used for the polymerase chain reaction were optimized so as to obtain more than 95% of PCR products giving 500bp-sequence reads.

The amplification products from genomic PCR using the oligonucleotides derived from the BAC subclones were subjected to automated dideoxy terminator sequencing reactions using a dye-primer cycle sequencing protocol. Following gel image analysis and DNA sequence extraction, sequence data were automatically processed with appropriate software to assess sequence quality and to detect the presence of biallelic sites among the pooled amplified fragments. Biallelic sites were systematically verified by comparing the sequences of both strands of each pool.

The detection limit for the frequency of biallelic polymorphisms detected by sequencing pools of 100 individuals is 0.3 +/- 0.05 for the minor allele, as verified by sequencing pools of known allelic frequencies. Thus, the biallelic markers selected by this method will be "informative biallelic markers" since they have a frequency of 0.3 to 0.5 for the minor allele and 0.5 to 0.7 for the major allele, therefore an average heterozygosity rate higher than 42%.

Example 3 describes the preparation of genomic DNA samples from the individuals screened

to identify biallelic markers.

Example 3

The population used in order to generate biallelic markers in the region of interest consisted of ca. 100 unrelated individuals corresponding to a French heterogeneous population.

5 DNA was extracted from peripheral venous blood of each donor as follows.

30 ml of blood were taken in the presence of EDTA. Cells (pellet) were collected after centrifugation for 10 minutes at 2000 rpm. Red cells were lysed by a lysis solution (50 ml final volume : 10 mM Tris pH7.6; 5 mM MgCl₂; 10 mM NaCl). The solution was centrifuged (10 minutes, 2000 rpm) as many times as necessary to eliminate the residual red cells present in the supernatant,
10 after resuspension of the pellet in the lysis solution.

The pellet of white cells was lysed overnight at 42°C with 3.7 ml of lysis solution composed of:

- 3 ml TE 10-2 (Tris-HCl 10 mM, EDTA 2 mM) / NaCl 0.4 M
- 200 µl SDS 10%
- 15 - 500 µl K-proteinase (2 mg K-proteinase in TE 10-2 / NaCl 0.4 M).

For the extraction of proteins, 1 ml saturated NaCl (6M) (1/3.5 v/v) was added. After vigorous agitation, the solution was centrifuged for 20 minutes at 10000 rpm.

For the precipitation of DNA, 2 to 3 volumes of 100% ethanol were added to the previous supernatant, and the solution was centrifuged for 30 minutes at 2000 rpm. The DNA solution was
20 rinsed three times with 70% ethanol to eliminate salts, and centrifuged for 20 minutes at 2000 rpm. The pellet was dried at 37°C, and resuspended in 1 ml TE 10-1 or 1 ml water. The DNA concentration was evaluated by measuring the OD at 260 nm (1 unit OD = 50 µg/ml DNA).

To determine the presence of proteins in the DNA solution, the OD 260 / OD 280 ratio was determined. Only DNA preparations having a OD 260 / OD 280 ratio between 1.8 and 2 were used in
25 the subsequent steps described below.

DNA Amplification

Once each BAC was isolated, pairs of primers, each one defining a 500 bp-amplification fragment, were designed. Each of the primers contained a common oligonucleotide tail upstream of the specific bases targeted for amplification, allowing the amplification products from each set of
30 primers to be sequenced using the common sequence as a sequencing primer. The primers used for the genomic amplification of sequences derived from BACs were defined with the OSP software (Hillier L. and Green P. Methods Appl., 1991, 1: 124-8). The synthesis of primers was performed following the phosphoramidite method, on a GENSET UFPS 24.1 synthesizer.

Example 4 provides the procedures used in the amplification reactions.

35

Example 4

The amplification of each sequence was performed by PCR (Polymerase Chain Reaction) as follows:

	- final volume	50 μ l
	- genomic DNA	100 ng
5	- $MgCl_2$	2 mM
	- dNTP (each)	200 μ M
	- primer (each)	7.5 pmoles
	- Ampli Taq Gold DNA polymerase (Perkin)	1 unit
	- PCR buffer (10X=0.1 M Tris HCl pH 8.3, 0.5 M KCl) 1X.	

10 The amplification was performed on a Perkin Elmer 9600 Thermocycler or MJ Research PTC200 with heating lid. After heating at 94°C for 10 minutes, 35 cycles were performed. Each cycle comprised: 30 sec at 94°C, 1 minute at 55°C, and 30 sec at 72°C. For final elongation, 7 minutes at 72°C ended the amplification.

The obtained quantity of amplification products was determined on 96-well microtiter plates, using a fluorimeter and Picogreen as intercalating agent (Molecular Probes).

The sequences of the amplification products were determined for each of the approximately 100 individuals from whom genomic DNA was obtained. Those amplification products which contained biallelic markers were identified.

Figure 1 shows the locations of the biallelic markers along the 8p23 BAC contig. This first set of markers corresponds to a medium density map of the candidate locus, with an inter-marker distance averaging 50kb-150kb.

A second set of biallelic markers was then generated as described above in order to provide a very high-density map of the region identified using the first set of markers which can be used to conduct association studies, as explained below. The high density map has markers spaced on average every 2-50kb.

The biallelic markers were then used in association studies as described below.

Collection of DNA samples from affected and non-affected individuals

Prostate cancer patients were recruited according to clinical inclusion criteria based on pathological or radical prostatectomy records. Control cases included in this study were both ethnically- and age-matched to the affected cases; they were checked for both the absence of all clinical and biological criteria defining the presence or the risk of prostate cancer, and for the absence of related familial prostate cancer cases. Both affected and control individuals corresponded to unrelated cases.

The two following pools of independent individuals were used in the association studies. The first pool, comprising individuals suffering from prostate cancer, contained 185 individuals. Of these

185 cases of prostate cancer, 45 cases were sporadic and 140 cases were familial. The second pool, the control pool, contained 104 non-diseased individuals.

Haplotype analysis was conducted using additional diseased (total samples: 281) and control samples (total samples: 130), from individuals recruited according to similar criteria.

5

Genotyping Affected and Control Individuals

The general strategy to perform the association studies was to individually scan the DNA samples from all individuals in each of the two populations described above in order to establish the allele frequencies of the above described biallelic markers in each of these populations.

Allelic frequencies of the above-described biallelic markers in each population were
10 determined by performing microsequencing reactions on amplified fragments obtained by genomic PCR performed on the DNA samples from each individual.

DNA samples and amplification products from genomic PCR were obtained in similar conditions as those described above for the generation of biallelic markers, and subjected to automated microsequencing reactions using fluorescent ddNTPs (specific fluorescence for each
15 ddNTP) and the appropriate oligonucleotide microsequencing primers which hybridized just upstream of the polymorphic base. Once specifically extended at the 3' end by a DNA polymerase using the complementary fluorescent dideoxynucleotide analog (thermal cycling), the primer was precipitated to remove the unincorporated fluorescent ddNTPs. The reaction products were analyzed by electrophoresis on ABI 377 sequencing machines.

20

Example 5 describes one microsequencing procedure.

Example 5

5 μ l of PCR products in a microtiter plate were added to 5 μ l purification mix {2U SAP (Amersham); 2U Exonuclease I (Amersham); 1 μ l SAP10X buffer: 400mM Tris-HCl pH8, 100 mM MgCl₂; H₂O final volume 5 μ l}. The reaction mixture was incubated 30 minutes at 37°C, and
25 denatured 10 minutes at 94°C. After 10 sec centrifugation, the microsequencing reaction was performed on line with the whole purified reaction mixture (10 μ l) in the microplate using 10 pmol microsequencing oligonucleotide (23mers, GENSET, crude synthesis, 5 OD), 0.5 U Thermosequenase (Amersham), 1.25 μ l Thermosequenase 16X buffer (Amersham), both of the fluorescent ddNTPs (Perkin Elmer) corresponding to the polymorphism {0.025 μ l ddTTP and ddCTP, 0.05 μ l ddATP and
30 ddGTP}, H₂O to a final volume of 20 μ l. A PCR program on a GeneAmp 9600 thermocycler was carried out as follows: 4 minutes at 94°C; 5 sec at 55°C / 10 sec at 94°C for 20 cycles. The reaction product was incubated at 4°C until precipitation. The microtiter plate was centrifuged 10 sec at 1500 rpm. 19 μ l MgCl₂ 2mM and 55 μ l 100 % ethanol were added in each well. After 15 minute incubation at room temperature, the microtiter plate was centrifuged at 3300 rpm 15 minutes at 4°C.
35 Supernatants were discarded by inverting the microtitre plate on a box folded to proper size and by centrifugation at 300 rpm 2 minutes at 4°C afterwards. The microplate was then dried 5 minutes in a

vacuum drier. The pellets were resuspended in 2.5 µl formamide EDTA loading buffer (0.7µl of 9 µg/µl dextran blue in 25 mM EDTA and 1.8 µl formamide). A 10% polyacrylamide gel / 12 cm / 64 wells was pre-run for 5 minutes on a 377 ABI 377 sequencer. After 5 minutes denaturation at 100°C, 0.8 µl of each microsequencing reaction product was loaded in each well of the gel. After migration
5 (2 h 30 for 2 microtiter plates of PCR products per gel), the fluorescent signals emitted by the incorporated ddNTPs were analyzed on the ABI 377 sequencer using the GENESCAN software (Perkin Elmer). Following gel analysis, data were automatically processed with a software that allowed the determination of the alleles of biallelic markers present in each amplified fragment.

I.D. Initial Association Studies

10 Association studies were run in two successive steps. In a first step, a rough localization of the candidate gene was achieved by determining the frequencies of the biallelic markers of Figure 1 in the affected and unaffected populations. The results of this rough localization are shown in Figure 2. This analysis indicated that a gene responsible for prostate cancer was located near the biallelic marker designated 4-67.

15 In a second phase of the analysis, the position of the gene responsible for prostate cancer was further refined using the very high density set of markers described above. The results of this localization are shown in Figure 3.

As shown in Figure 3, the second phase of the analysis confirmed that the gene responsible for prostate cancer was near the biallelic marker designated 4-67, most probably within a ca. 150kb
20 region comprising the marker.

Haplotype analysis

The allelic frequencies of each of the alleles of biallelic markers 99-123, 4-26, 4-14, 4-77, 99-217, 4-67, 99-213, 99-221, and 99-135 (SEQ ID NOs: 21-38) were determined in the affected and unaffected populations. Table 1 lists the internal identification numbers of the markers used in the
25 haplotype analysis (SEQ ID NOs: 21-38), the alleles of each marker, the most frequent allele in both unaffected individuals and individuals suffering from prostate cancer, the least frequent allele in both unaffected individuals and individuals suffering from prostate cancer, and the frequencies of these alleles in each population.

Among all the theoretical potential different haplotypes based on 2 to 9 markers, 11
30 haplotypes showing a strong association with prostate cancer were selected. The results of these haplotype analyses are shown in Figure 4.

Figures 2, 3, and 4 aggregate linkage analysis results with sequencing results which permitted the physical order and/or the distance between markers to be estimated.

The significance of the values obtained in Figure 4 are underscored by the following results
35 of computer simulations. For the computer simulations, the data from the affected individuals and the unaffected controls were pooled and randomly allocated to two groups which contained the same

number of individuals as the affected and unaffected groups used to compile the data summarized in Figure 4. A haplotype analysis was run on these artificial groups for the six markers included in haplotype 5 of Figure 4. This experiment was reiterated 100 times and the results are shown in Figure 5. Among 100 iterations, only 5% of the obtained haplotypes are present with a p-value below 1×10^{-4} as compared to the p-value of 9×10^{-7} for haplotype 5 of Figure 4. Furthermore, for haplotype 5 of Figure 4, only 6% of the obtained haplotypes have a significance level below 5×10^{-3} , while none of them show a significance level below 5×10^{-5} .

Thus, using the data of Figure 4 and evaluating the associations for single marker alleles or for haplotypes will permit estimation of the risk a corresponding carrier has to develop prostate cancer. Significance thresholds of relative risks will be adapted to the reference sample population used.

The diagnostic techniques may employ a variety of methodologies to determine whether a test subject has a biallelic marker pattern associated with an increased risk of developing prostate cancer or suffers from prostate cancer resulting from a mutant PG1 allele. These include any method enabling the analysis of individual chromosomes for haplotyping, such as family studies, single sperm DNA analysis or somatic hybrids.

In each of these methods, a nucleic acid sample is obtained from the test subject and the biallelic marker pattern for one or more of the biallelic markers listed in Figures 4, 6A and 6B is determined. The biallelic markers listed in Figure 6A are those which were used in the haplotype analysis of Figure 4. The first column of Figure 6A lists the BAC clones in which the biallelic markers lie. The second column of Figure 6A lists the internal identification number of the marker. The third column of Figure 6A lists the sequence identification number for a first allele of the biallelic markers. The fourth column of Figure 6A lists the sequence identification number for a second allele of the biallelic markers. For example, the first allele of the biallelic marker 99-123 has the sequence of SEQ ID NO:21 and the second allele of the biallelic marker has the sequence of SEQ ID NO: 30.

The fifth column of Figure 6A lists the sequences of upstream primers which is used to generate amplification products containing the polymorphic bases of the biallelic markers. The sixth column of Figure 6A lists the sequence identification numbers for the upstream primers.

The seventh column of Figure 6A lists the sequences of downstream primers which is used to generate amplification products containing the polymorphic bases of the biallelic markers. The eighth column of Figure 6A lists the sequence identification numbers for the downstream primers.

The ninth column of Figure 6A lists the position of the polymorphic base in the amplification products generated using the upstream and downstream primers. The tenth column lists the identities of the polymorphic bases found at the polymorphic positions in the biallelic markers. The eleventh and twelfth columns list the locations of microsequencing primers in the biallelic markers which can be used to determine the identities of the polymorphic bases.

In addition to the biallelic markers of SEQ ID NOs: 21-38, other biallelic markers (designated 99-1482, 4-73, 4-65) have been identified which are closely linked to one or more of the biallelic markers of SEQ ID NOs: 21-38, SEQ ID NOs: 57-62, and the PG1 gene. These biallelic markers include the markers of SEQ ID NOs: 57-62, which are listed in Figure 6B. The columns in Figure 6B
5 are identical to the corresponding columns in Figure 6A. SEQ ID NOs: 58, 59, 61, and 62 lie within the PG1 gene of SEQ ID NO:1 at the positions indicated in the accompanying Sequence Listing.

Genetic analysis of these additional biallelic markers is performed as follows. Nucleic acid samples are obtained from individuals suffering from prostate cancer and unaffected individuals. The frequencies at which each of the two alleles occur in the affected and unaffected populations is
10 determined using the methodologies described above. Association values are calculated to determine the correlation between the presence of a particular allele or spectrum of alleles and prostate cancer. The markers of SEQ ID NOs: 21-38 may also be included in the analysis used to calculate the risk factors. The markers of SEQ ID NOs: 21-38 and SEQ ID NOs: 57-62 is used in diagnostic techniques, such as those described below, to determine whether an individual is at risk for
15 developing prostate cancer or suffers from prostate cancer as a result of a mutation in the PG1 gene.

Example 6 describes methods for determining the biallelic marker pattern.

Example 6

A nucleic acid sample is obtained from an individual to be tested for susceptibility to prostate cancer or PG1 mediated prostate cancer. The nucleic acid sample is an RNA sample or a DNA
20 sample.

A PCR amplification is conducted using primer pairs which generate amplification products containing the polymorphic nucleotides of one or more biallelic markers associated with prostate cancer-related forms of PG1, such as the biallelic markers of SEQ ID NOs: 21-38, SEQ ID NOs: 57-62, biallelic markers which are in linkage disequilibrium with the biallelic markers of SEQ ID NOs:
25 21-38, SEQ ID NOs: 57-62, biallelic markers in linkage disequilibrium with the PG1 gene, or combinations thereof. In some embodiments, the PCR amplification is conducted using primer pairs which generate amplification products containing the polymorphic nucleotides of several biallelic markers. For example, in one embodiment, amplification products containing the polymorphic bases of several biallelic markers selected from the group consisting of SEQ ID NOs: 21-38, SEQ ID NOs:
30 57-62, and biallelic markers which are in linkage disequilibrium with the biallelic markers of SEQ ID NOs: 21-38, SEQ ID NOs: 57-62 or with the PG1 gene is generated. In another embodiment, amplification products containing the polymorphic bases of two or more biallelic markers selected from the group consisting of SEQ ID NOs: 21-38, SEQ ID NOs: 57-62, and biallelic markers which are in linkage disequilibrium with the biallelic markers of SEQ ID NOs: 21-38, SEQ ID NOs: 57-62
35 or with the PG1 gene is generated. In another embodiment, amplification products containing the polymorphic bases of five or more biallelic markers selected from the group consisting of SEQ ID

NOs: 21-38, SEQ ID NOs: 57-62, and biallelic markers which are in linkage disequilibrium with the biallelic markers of SEQ ID NOs: 21-38, SEQ ID NOs: 57-62 or with the PG1 gene is generated. In another embodiment, amplification products containing the polymorphic bases of more than five of the biallelic markers selected from the group consisting of SEQ ID NOs: 21-38, SEQ ID NOs: 57-62, and biallelic markers which are in linkage disequilibrium with the biallelic markers of SEQ ID NOs: 21-38, SEQ ID NOs: 57-62 or with the PG1 gene is generated.

For example, the primers used to generate the amplification products may comprise the primers listed in Figure 6A or 6B (SEQ ID NOs: 39-56 and SEQ ID NOs: 63-68). Figures 6A and Figure 6B provide exemplary primers which is used in the amplification reactions and the identities and locations of the polymorphic bases in the amplification products which are produced with the exemplary primers. The sequences of each of the alleles of the biallelic markers resulting from amplification using the primers in Figures 6A and 6B are listed in the accompanying Sequence Listing as SEQ ID NOs: 21-38 and 57-62.

The PCR primers is oligonucleotides of 10, 15, 20 or more bases in length which enable the amplification of the polymorphic site in the markers. In some embodiments, the amplification product produced using these primers is at least 100 bases in length (i.e. 50 nucleotides on each side of the polymorphic base). In other embodiments, the amplification product produced using these primers is at least 500 bases in length (i.e. 250 nucleotides on each side of the polymorphic base). In still further embodiments, the amplification product produced using these primers is at least 1000 bases in length (i.e. 500 nucleotides on each side of the polymorphic base).

It will be appreciated that the primers listed in Figure 6A and 6B are merely exemplary and that any other set of primers which produce amplification products containing the polymorphic nucleotides of one or more of the biallelic markers of SEQ ID NOs. 21-38 and SEQ ID NOs: 57-62 or biallelic markers in linkage disequilibrium with the sequences of SEQ ID NOs. 21-38 and SEQ ID NOs: 57-62 or with the PG1 gene, or a combination thereof is used in the diagnostic methods.

Following the PCR amplification, the identities of the polymorphic bases of one or more of the biallelic markers of SEQ ID NOs: 21-38 and SEQ ID NOs: 57-62, or biallelic markers in linkage disequilibrium with the sequences of SEQ ID NOs. 21-38 and SEQ ID NOs: 57-62 or with the PG1 gene, or a combination thereof, are determined. The identities of the polymorphic bases is determined using the microsequencing procedures described in Example 5 above and the microsequencing primers listed as features in the sequences of SEQ ID NOs: 21-38 and SEQ ID NOs: 57-62. It will be appreciated that the microsequencing primers listed as features in the sequences of SEQ ID NOs: 21-38 and SEQ ID NOs: 57-62 are merely exemplary and that any primer having a 3' end near the polymorphic nucleotide, and preferably immediately adjacent to the polymorphic nucleotide, is used. Alternatively, the microsequencing analysis is performed as described in Pastinen et al., Genome Research 7:606-614 (1997), which is described in more detail below.

Alternatively, the PCR product is completely sequenced to determine the identities of the polymorphic bases in the biallelic markers. In another method, the identities of the polymorphic bases in the biallelic markers is determined by hybridizing the amplification products to microarrays containing allele specific oligonucleotides specific for the polymorphic bases in the biallelic markers.

5 The use of microarrays comprising allele specific oligonucleotides is described in more detail below.

It will be appreciated that the identities of the polymorphic bases in the biallelic markers is determined using techniques other than those listed above, such as conventional dot blot analyses.

Nucleic acids used in the above diagnostic procedures may comprise at least 10 consecutive nucleotides in the biallelic markers of SEQ ID NOs: 21-38 and SEQ ID NOs: 57-62 or the sequences
10 complementary thereto. Alternatively, the nucleic acids used in the above diagnostic procedures may comprise at least 15 consecutive nucleotides in the biallelic markers of SEQ ID NOs: 21-38 and SEQ ID NOs: 57-62 or the sequences complementary thereto. In some embodiments, the nucleic acids used in the above diagnostic procedures may comprise at least 20 consecutive nucleotides in the biallelic markers of SEQ ID NOs: 21-38 and SEQ ID NOs: 57-62 or the sequences complementary thereto. In
15 still other embodiments, the nucleic acids used in the above diagnostic procedures may comprise at least 30 consecutive nucleotides in the biallelic markers of SEQ ID NOs: 21-38 and SEQ ID NOs: 57-62 or the sequences complementary thereto. In further embodiments, the nucleic acids used in the above diagnostic procedures may comprise more than 30 consecutive nucleotides in the biallelic markers of SEQ ID NOs: 21-38 and SEQ ID NOs: 57-62 or the sequences complementary thereto. In
20 still further embodiments, the nucleic acids used in the above diagnostic procedures may comprise the entire sequence of the biallelic markers of SEQ ID NOs: 21-38 and SEQ ID NOs: 57-62 or the sequences complementary thereto.

I.E. Identification and Sequencing, of the PG1 Gene, and Localization of the PG1 Protein

The above haplotype analysis indicated that 171kb of genomic DNA between biallelic
25 markers 4-14 and 99-221 totally or partially contains a gene responsible for prostate cancer. Therefore, the protein coding sequences lying within this region were characterized to locate the gene associated with prostate cancer. This analysis, described in further detail below, revealed a single protein coding sequence in the 171 kb, which was designated as the PG1 gene.

Template DNA for sequencing the PG1 gene was obtained as follows. BACs 189EO8 and
30 463FO1 were subcloned as previously described. Plasmid inserts were first amplified by PCR on PE 9600 thermocyclers (Perkin-Elmer), using appropriate primers, AmpliTaqGold (Perkin-Elmer), dNTPs (Boehringer), buffer and cycling conditions as recommended by the Perkin-Elmer Corporation.

PCR products were then sequenced using automatic ABI Prism 377 sequencers (Perkin Elmer, Applied Biosystems Division, Foster City, CA). Sequencing reactions were performed using PE 9600
35 thermocyclers (Perkin Elmer) with standard dye-primer chemistry and ThermoSequenase (Amersham Life Science). The primers were labeled with the JOE, FAM, ROX and TAMRA dyes. The dNTPs and

ddNTPs used in the sequencing reactions were purchased from Boehringer. Sequencing buffer, reagent concentrations and cycling conditions were as recommended by Amersham.

Following the sequencing reaction, the samples were precipitated with EtOH, resuspended in formamide loading buffer, and loaded on a standard 4% acrylamide gel. Electrophoresis was performed
5 for 2.5 hours at 3000V on an ABI 377 sequencer, and the sequence data were collected and analyzed using the ABI Prism DNA Sequencing Analysis Software, version 2.1.2.

The sequence data obtained as described above were transferred to a proprietary database, where quality control and validation steps were performed. A proprietary base-caller ("Trace"), working using a Unix system automatically flagged suspect peaks, taking into account the shape of the peaks, the inter-
10 peak resolution, and the noise level. The proprietary base-caller also performed an automatic trimming. Any stretch of 25 or fewer bases having more than 4 suspect peaks was considered unreliable and was discarded. Sequences corresponding to cloning vector oligonucleotides were automatically removed from the sequence. However, the resulting sequence may contain 1 to 5 bases belonging to the vector sequences at their 5' end. If needed, these can easily be removed on a case by case basis.

15 The genomic sequence of the PG1 gene is provided in the accompanying Sequence Listing and is designated as SEQ ID NO: 1.

Potential exons in BAC-derived human genomic sequences were located by homology searches on protein, nucleic acid and EST (Expressed Sequence Tags) public databases. Main public databases were locally reconstructed. The protein database, NRPU (Non-redundant Protein Unique) is formed by a
20 non-redundant fusion of the Genpept (Benson D.A. et al., Nucleic Acids Res. 24: 1-5 (1996), Swissprot (Bairoch, A. and Apweiler, R., Nucleic Acids Res. 24: 21-25 (1996) and PIR/NBRF (George, D.G. et al., Nucleic Acids Res. 24:17-20 (1996) databases. Redundant data were eliminated by using the NRDB software (Benson et al., supra) and internal repeats were masked with the XNU software (Benson et al., supra). Homologies found using the NRPU database allowed the identification of sequences
25 corresponding to potential coding exons related to known proteins.

The EST local database is composed by the gbest section (1-9) of GenBank (Benson et al., supra), and thus contains all publicly available transcript fragments. Homologies found with this database allowed the localization of potentially transcribed regions.

The local nucleic acid database contained all sections of GenBank and EMBL (Rodriguez-
30 Tome, P. et al., Nucleic Acids Res. 24: 6-12 (1996) except the EST sections. Redundant data were eliminated as previously described.

Similarity searches in protein or nucleic acid databases were performed using the BLAS software (Altschul, S.F. et al., J. Mol. Biol. 215: 403-410 (1990). Alignments were refined using the Fasta software, and multiple alignments used Clustal W. Homology thresholds were adjusted for each
35 analysis based on the length and the complexity of the tested region, as well as on the size of the reference database.

Potential exon sequences identified as above were used as probes to screen cDNA libraries. Extremities of positive clones were sequenced and the sequence stretches were positioned on the genomic sequence of SEQ ID NO:1. Primers were then designed using the results from these alignments in order to enable the PG1 cloning procedure described below.

5

Cloning PG1 cDNA

PG1 cDNA was obtained as follows. 4:1 of ethanol suspension containing 1mg of human prostate total RNA (Clontech laboratories, Inc., Palo Alto, USA; catalogue N. 64038-1, lot 7040869) was centrifuged, and the resulting pellet was air dried for 30 minutes at room temperature.

First strand cDNA synthesis was performed using the AdvantageTM RT-for-PCR kit (Clontech laboratories, Inc., Palo Alto, USA; catalogue N. K1402-1). 1:1 of 20mM solution of primer PGRT32: TTTTTTTTTTTTTTTTTTGGAAAT (SEQ ID NO:10) was added to 12.5 :l of RNA solution in water, heated at 74°C for two and a half minutes and rapidly quenched in an ice bath. 10:1 of 5xRT buffer (50mM Tris-HCl, pH 8.3, 75mM KCl, 3 mM MgCl₂), 2.5 :l of dNTP mix (10mM each), 1.25:1 of human recombinant placental RNA inhibitor were mixed with 1 ml of MMLV reverse transcriptase (200 units). 6.5:1 of this solution were added to RNA-primer mix and incubated at 42°C for one hour. 80:1 of water were added and the solution was incubated at 94°C for 5 minutes.

5:1 of the resulting solution were used in a Long Range PCR reaction with hot start, in 50 :l final volume, using 2 units of rTHXL, 20 pmol/ μ l of each of GC1.5p.1: CTGTCCCTGGTGCTCCACACGTACTC (SEQ ID NO:6) or GC1.5p2: TGGTGCTCCACACGTACTCCATGCGC (SEQ ID NO: 7) and GC1.3p: CTTGCCTGCTGGAGACACAGAATTCGATAGCAC (SEQ ID NO:9) primers with 35 cycles of elongation for 6 minutes at 67°C in thermocycler.

The sequence of the PG1 cDNA obtained as described above (SEQ ID NO 3) is provided in the accompanying Sequence Listing. Results of Northern blot analysis of prostate mRNAs support the existence of a major PG1 cDNA having a 5-6kb length.

Characterization of the PG1 Gene

The intron/exon structure of the gene was deduced by aligning the mRNA sequence from the cDNA of SEQ ID NO:3 and the genomic DNA sequence of SEQ ID NO: 1.

The positions of the introns and exons in the PG1 genomic DNA are provided in Figures 7 and 8. Figure 7 lists positions of the start and end nucleotides defining each of the at least 8 exons (labeled Exons A-H) in the sequence of SEQ ID NO: 1, the locations and phases of the 5' and 3' splice sites in the sequence of SEQ ID NO: 1, the position of the stop codon in the sequence of SEQ ID NO: 1, and the position of the polyadenylation site in the sequence of SEQ ID NO: 1. Figure 8 shows the positions of the exons within the PG1 genomic DNA and the PG1 mRNA, the location of a tyrosine phosphatase retro-pseudogene in the PG1 genomic DNA, the positions of the coding region in the mRNA, and the locations of the polyadenylation signal and polyA stretch in the mRNA.

As indicated in Figures 7 and 8, the PG1 gene comprises at least 8 exons, and spans more than 52kb. The first intron contains a tyrosine phosphatase retropseudogene. A G/C rich putative promoter region lies between nucleotide 1629 and 1870 of SEQ ID NO: 1. A CCAAT box is present at nucleotide 1661 of SEQ ID NO: 1. The promoter region was identified as described in Prestridge, 5 D.S., Predicting Pol II Promoter Sequences Using Transcription Factor Binding Sites, J. Mol. Biol. 249:923-932 (1995).

It is possible that the methionine listed as being the initiating methionine in the PG1 protein sequence of SEQ ID NO: 4 (based on the cDNA sequence of SEQ ID NO: 3) may actually be downstream but in phase with another methionine which acts as the initiating methionine. The 10 genomic DNA sequence of SEQ ID NO: 1 contains a methionine upstream from the methionine at position number 1 of the protein sequence of SEQ ID NO: 4. If the upstream methionine is in fact the authentic initiation site, the sequence of the PG1 protein would be that of SEQ ID NO: 5. This possibility is investigated by determining the exact position of the 5N end of the PG1 mRNA as follows.

15 One way to determine the exact position of the 5N end of the PG1 mRNA is to perform a 5NRACE reaction using the Marathon-Ready human prostate cDNA kit from Clontech (Catalog. No. PT1156-1). For example, the RACE reaction may employ the PG1 primers PG15RACE196 CAATATCTGGACCCCGGTGTAATTCTC (SEQ ID NO: 8) as the first primer. The second primer in the RACE reaction is PG15RACE130n having the sequence 20 GGTCGTCCAGCGCTTGGTAGAAG (SEQ ID NO: 2). The sequence analysis of the resulting PCR product, or the product obtained with other PG1 specific primers, will give the exact sequence of the initiation point of the PG1 transcript.

Alternatively, the 5Nsequence of the PG1 transcript can be determined by conducting a PCR amplification with a series of primers extending from the 5Nend of the presently identified coding 25 region. In any event, the present invention contemplates use of PG1 nucleic acids and/or polypeptides coding for or corresponding to either SEQ ID NO:4 or SEQ ID NO:5 or fragments thereof.

It is also possible that alternative splicing of the PG1 gene may result in additional translation products not described above. It is also possible that there are sequences upstream or downstream of the genomic sequence of SEQ ID NO: 1 which contribute to the translation products of the gene. 30 Finally, alternative promoters may result in PG1derived transcripts other than those described herein.

The promoter activity of the region between nucleotides 1629 and 1870 can be verified as described below. Alternatively, should this region lack promoter activity, the promoter responsible for driving expression of the PG1 gene is identified as described below.

Genomic sequences lying upstream of the PG1 gene are cloned into a suitable promoter reporter 35 vector, such as the pSEAP-Basic, pSEAP-Enhancer, p β gal-Basic, p β gal-Enhancer, or pEGFP-1 Promoter Reporter vectors available from Clontech. Briefly, each of these promoter reporter vectors

include multiple cloning sites positioned upstream of a reporter gene encoding a readily assayable protein such as secreted alkaline phosphatase, β galactosidase, or green fluorescent protein. The sequences upstream of the PG1 coding region are inserted into the cloning sites upstream of the reporter gene in both orientations and introduced into an appropriate host cell. The level of reporter protein is
5 assayed and compared to the level obtained from a vector which lacks an insert in the cloning site. The presence of an elevated expression level in the vector containing the insert with respect to the control vector indicates the presence of a promoter in the insert. If necessary, the upstream sequences can be cloned into vectors which contain an enhancer for augmenting transcription levels from weak promoter sequences. A significant level of expression above that observed with the vector lacking an insert
10 indicates that a promoter sequence is present in the inserted upstream sequence.

Promoter sequences within the upstream genomic DNA is further defined by constructing nested deletions in the upstream DNA using conventional techniques such as Exonuclease III digestion. The resulting deletion fragments can be inserted into the promoter reporter vector to determine whether the deletion has reduced or obliterated promoter activity. In this way, the boundaries of the promoters is
15 defined. If desired, potential individual regulatory sites within the promoter is identified using site directed mutagenesis or linker scanning to obliterate potential transcription factor binding sites within the promoter individually or in combination. The effects of these mutations on transcription levels is determined by inserting the mutations into the cloning sites in the promoter reporter vectors.

Sequences within the PG1 promoter region which are likely to bind transcription factors is
20 identified by homology to known transcription factor binding sites or through conventional mutagenesis or deletion analyses of reporter plasmids containing the promoter sequence. For example, deletions is made in a reporter plasmid containing the promoter sequence of interest operably linked to an assayable reporter gene. The reporter plasmids carrying various deletions within the promoter region are transfected into an appropriate host cell and the effects of the deletions on expression levels is assessed.
25 Transcription factor binding sites within the regions in which deletions reduce expression levels is further localized using site directed mutagenesis, linker scanning analysis, or other techniques familiar to those skilled in the art.

The promoters and other regulatory sequences located upstream of the PG1 gene is used to design expression vectors capable of directing the expression of an inserted gene in a desired spatial,
30 temporal, developmental, or quantitative manner. For example, since the PG1 promoter is presumably active in the prostate, it can be used to construct expression vectors for directing gene expression in the prostate.

Preferably, in such expression vectors, the PG1 promoter is placed near multiple restriction sites to facilitate the cloning of an insert encoding a protein for which expression is desired downstream of the
35 promoter, such that the promoter is able to drive expression of the inserted gene. The promoter is inserted in conventional nucleic acid backbones designed for extrachromosomal replication, integration

into the host chromosomes or transient expression. Suitable backbones for the present expression vectors include retroviral backbones, backbones from eukaryotic episomes such as SV40 or Bovine Papilloma Virus, backbones from bacterial episomes, or artificial chromosomes.

Preferably, the expression vectors also include a polyA signal downstream of the multiple
5 restriction sites for directing the polyadenylation of mRNA transcribed from the gene inserted into the expression vector.

Nucleic acids encoding proteins which interact with sequences in the PG1 promoter is identified using one-hybrid systems such as those described in the manual accompanying the Matchmaker One-Hybrid System kit available from Clontech (Catalog No. K1603-1). Briefly, the Matchmaker One-
10 hybrid system is used as follows. The target sequence for which it is desired to identify binding proteins is cloned upstream of a selectable reporter gene and integrated into the yeast genome. Preferably, multiple copies of the target sequences are inserted into the reporter plasmid in tandem.

A library comprised of fusions between cDNAs to be evaluated for the ability to bind to the promoter and the activation domain of a yeast transcription factor, such as GAL4, is transformed into the
15 yeast strain containing the integrated reporter sequence. The yeast are plated on selective media to select cells expressing the selectable marker linked to the promoter sequence. The colonies which grow on the selective media contain genes encoding proteins which bind the target sequence. The inserts in the genes encoding the fusion proteins are further characterized by sequencing. In addition, the inserts is inserted into expression vectors or in vitro transcription vectors. Binding of the polypeptides encoded by the
20 inserts to the promoter DNA is confirmed by techniques familiar to those skilled in the art, such as gel shift analysis or DNase protection analysis.

Analysis of PG1 Protein Sequence

The PG1 cDNA of SEQ ID NO: 3 encodes a 353 amino-acid protein (SEQ ID NO:4). As indicated in the accompanying Sequence Listing, a Prosite analysis indicated that the PG1 protein has
25 a leucine zipper motif, a potential glycosylation site, 3 potential casein kinase II phosphorylation sites, a potential cAMP dependent protein kinase phosphorylation site, 2 potential tyrosine kinase phosphorylation sites, 4 potential protein kinase C phosphorylation sites, 5 potential N-myristoylation sites, 1 potential tyrosine sulfation site, and one potential amidation site.

A search for membrane associated domains was conducted according to the methods
30 described in Argos, P. et al., Structural Prediction of Membrane-bound Proteins, *Elur. J. Biochem.* 128:565-575 (1982); Klein et al., *Biochimica & Biophysica Acta* 815:468-476 (1985); and Eisenberg et al., *J. Mol. Biol.* 179:125-142 (1984). The search revealed 5 potential transmembrane domains predicted to be integral membrane domains. These results suggest that the PG1 protein is likely to be membrane-associated and is an integral membrane protein.

35 A homology search was conducted to identify proteins homologous to the PG1 protein. Several proteins were identified which share homology with the PG1 protein. Figure 9 lists the

accession numbers of several proteins which share homology with the PG1 protein in three regions designated box1, box2 and box3.

It will be appreciated that each of the motifs described above is also present in the protein of SEQ ID NO: 5, which would be produced if by translation initiation translated from the potential upstream methionine in the nucleic acid of SEQ ID NO: 1.

As indicated in Figure 9, a distinctive pattern of homology to box 1, box 2 (SEQ ID NOs: 11-14) and box 3 (SEQ ID NOs: 15-20) is found amongst acyl glycerol transferases. For example, the *plsC* protein from *E. coli* (Accession Number P26647) shares homology with the box1 and box2 sequences, but not the box 3 sequence, of the PG1 protein. The product of this gene transfers acyl from acyl-coenzymeA to the sn2 position of 1-Acyl-sn-glycerol-3-phosphate (lysophosphatidic acid, LPA)(Coleman J., Mol Gen Genet. 1992 Mar 1; 232(2): 295-303).

Box1 and box2 homologies, but not box 3 homologies, are also found in the SLCI gene product from baker's yeast (Accession Number P33333) and the mouse gene AB005623. Each of these genes are able to complement *in vivo* mutations in the bacterial *plsC* gene. (Nagiec MM, Wells GB, Lester RL, Dickson RC, J. Biol. Chem., 1993 Oct 15; 268(29): 22156-22163, A suppressor gene that enables *Saccharomyces cerevisiae* to grow without making sphingolipids encodes a protein that resembles an *Escherichia coli* fatty acyltransferase; and Kume K, Shimizu T, Biochem. Biophys. Res. Commun. 1997, Aug. 28; 237(3): 663-666, cDNA cloning and expression of murine 1-acyl-sn-glycerol-3-phosphate acyltransferase).

Recently two different human homologues of the mouse AB005623 gene, Accession Numbers U89336 and U56417 were cloned and found to be localized to human chromosomes 6 and 9 (Eberhardt. C., Gray, P.W. and Tjoelker, L.W., J. Biol. Chem. 1997; 272, 20299-20305, Human lysophosphatidic acid acyltransferase cDNA cloning, expression, and localization to chromosome 9q34.3; and West, J., Tompkins, C.K., Balantac, N., Nudelman, E., Meengs, B., White, T., Bursten, S., Coleman, J., Kumar, A., Singer, J.W. and Leung, D.W, DNA Cell Biol. 6, 691-701 (1997), Cloning and expression of two human lysophosphatidic acid acyltransferase cDNAs that enhance cytokine induced signaling responses in cells).

The enzymatic acylation of LPA results in 1,2-diacyl-sn-glycerol 3-phosphate, an intermediate to the biosynthesis of both glycerophospholipids and triacylglycerol. Several important signaling messengers participating in the transduction of mitogenic signals, induction of apoptosis, transmission of nerve impulses and other cellular responses mediated by membrane bound receptors belong to this metabolic pathway.

LPA itself is a potent regulator of mammalian cell proliferation. In fact, LPA is one of the major mitogens found in blood serum. (For a review: Durieux ME, Lynch KR, Trends Pharmacol. Sci. 1993 Jun; 14(6):249-254, Signaling properties of lysophosphatidic acid. LPA can act as a survival factor to inhibit apoptosis of primary cells; and Levine JS, Koh JS, Triaca V, Lieberthal W,

Am. J. Physiol. 1997 Oct; 273(4Pt2): F575-F585, Lysophosphatidic acid: a novel growth and survival factor for renal proximal tubular cells). This function of LPA is mediated by the lipid kinase phosphatidylinositol 3-kinase.

Phosphatidylinositol and its derivatives present another class of messengers emerging from the 1-acyl-sn-glycerol-3-phosphate acyltransferase pathway. (Toker A, Cantley LC, Nature 1997 Jun 12; 387(6634): 673-676, Signaling through the lipid products of phosphoinositide-3-OH kinase; Martin TF, Curr. Opin. Neurobiol. 1997 Jun; 7(3):331-338, Phosphoinositides as spatial regulators of membrane traffic; and Hsuan JJ, et al., Int. J. Biochem. Cell Biol. 1997 Mar 1st; 29(3): 415-435, Growth factor-dependent phosphoinositide signaling).

10 Cell growth, differentiation and apoptosis can be affected and modified by enzymes involved in this metabolic pathway. Consequently, alteration of this pathway could facilitate cancer cell progression. Modulation of the activity of enzymes in this pathway using agents such as enzymatic inhibitors could be a way to restore a normal phenotype to cancerous cells.

Ashagbley A, Samadder P, Bittman R, Erukulla RK, Byun HS, Arthur G have recently shown 15 that ether-linked analogue of lysophosphatidic acid: 4-O-hexadecyl-3(S)-O-methoxybutanephosphonate can effectively inhibit the proliferation of several human cancerous cell lines, including DU145 line of prostate cancer origin. (Anticancer Res 1996 Jul; 16(4A): 1813-1818, Synthesis of ether-linked analogues of lysophosphatidate and their effect on the proliferation of human epithelial cancer cells in vitro).

20 Structural differences between the PG1 family of cellular proteins and the functionally confirmed 1-acyl-sn-glycerol-3-phosphate acyltransferase family, evidenced by the existence of a different pattern of homology to box3, could point to unique substrate specificity in the phospholipid metabolic pathway, to specific interaction with other cellular components or to both.

Further analysis of the function of the PG1 gene can be conducted, for example, by 25 constructing knockout mutations in the yeast homologues of the PG1 gene in order to elucidate the potential function of this protein family, and to test potential substrate analogs in order to revert the malignant phenotype of human prostate cancer cells as described in Section VIII, below.

Example 7

Analysis of the Intracellular Localisation of the PG1 Isoforms

30 To study the intracellular localisation of PG1 protein, different isoforms of PG1 were cloned in the expression vector pEGFP-N1(Clontech), transfected and expressed in normal (PNT2A) or adenocarcinoma (PC3) prostatic cell line.

First, to generate cDNA inserts, 5' and 3' primers were synthesised allowing to amplify different regions of the PG1 open reading frame. Respectively, these primers were designed with an 35 internal EcoRI or BamHI site which allowed the insertion of the amplified product into the EcoRI and BamHI sites of the expression vector. The restriction sites were introduced into the primer so that

after cloning into pEGFP-N1, the PG1 open reading frame would be fused in frame, to the EGFP open reading frame. The translated protein would be a fusion between PG1 and EGFP. EGFP being a variant form of the GFP protein (Green Fluorescent Protein), it is possible to detect the intracellular localisation of the different PG1 isoforms by examining the fluorescence emitted by the EGFP fused protein.

The different forms that were analysed correspond either to different messengers identified by RT-PCR performed on total normal human prostatic RNA or to a truncated form resulting from a non sense mutation identified in a tumoural prostatic cell line LNCaP. The different PG1 constructions were transfected using the lipofectine technique and EGFP expression was examined 20 hours post transfection.

Name and description of the different forms transfected are listed below :

- A) PG1 includes all the coding exons from exon 1 to 8.
- B) PG1/1-4 corresponds to an alternative messenger which is due to an alternative splicing, joining exon 1 to exon 4, and resulting in the absence of exons 2 and 3.
- 15 C) PG1/1-5 corresponds to an alternative messenger which is due to an alternative splicing, joining exon 1 to exon 5, and resulting in the absence of exons 2, 3 and 4.
- D) PG1/1-7 includes exons 1 to 6, and corresponds to the mutated form identified in genomic DNA of the prostatic tumoural cell line LNCaP.

Cloning of the PG1 cDNA inserts in the EGFP-N1 expression vector

- 20 cDNAs from human prostate were obtained by RT-PCR using the Advantage RT-for-PCR Kit (CLONTECH ref K1402-2). First, 1µl of oligodT-containing PG1 specific primer PGRT32 TTTTTTTTTTTTTTTTGAAT (20pmoles) and 11.5 µl of DEPC treated H₂O were added to 1µl of total mRNA (1µg) extracted from human prostate (CLONTECH ref 64038-1). The mRNA was heat denaturated for 2.5 min at 74°C and then quickly chilled on ice. A mix containing 4µl of 5X
- 25 buffer, 1µl of dNTPs (10mM each), 0.5µl of recombinant RNase inhibitor (20U) and 1µl of MoMuLV Reverse Transcriptase (200U) was added to the denaturated mRNA. Reverse transcription was performed for 60 min. at 42°C. Enzymes were heat denaturated for 5 min. at 94°C. Then, 80µl of DEPC treated H₂O were added to the reaction mix and the cDNA mix was stored at -20°C. Primers PG15Eco3 (5' CCTGAATTCGCCGAGCTGAGAAGATGC 3'), and PG13Bam2 (5'
- 30 CCTGGATCCGCTTTAATAGTAACCCACAGGCAG 3') were used for PCR amplification of the different PG1 cDNAs. A 50µl PCR reaction mix containing 5µl of the previously prepared prostate cDNA mix, 15µl of 3.3X PCR buffer, 4µl of dNTPs (2.5mM each), 20pmoles of primer PG15Eco3, 20pmoles of primer PG13Bam2, 1µl of RthXL enzyme, 2.2µl Mg(OAc)₂ (Hot Start) was set up and amplification was performed for 35 cycles of 30 sec at 94°C, 10 min. at 72°C, 4 min. at 67°C after an
- 35 initial denaturation step of 10 min. at 94°C. Size and integrity of the PCR product was assessed by migration on a 1% agarose gel. 2µg of the amplification product were digested with 2.4 units of

EcoRI (PROMEGA ref R601A) and 2.0 units of BamHI (PROMEGA ref R602A) in 50µl of 1X Multicore buffer for 2 hours at 37°C. Enzymes were then heat inactivated for 20 min, at 68°C, DNA was phenol/chloroform extracted and ethanol-precipitated and its concentration was estimated by migration on a 1% agarose gel.

- 5 To prepare the vector, 2µg of pEGFP-N1 vector (CLONTECH ref 6085-1) were digested with 2.4 units of EcoRI (PROMEGA ref R601A) and 2.0 units of BamHI (PROMEGA ref R602A) in 50µl of 1X multicore buffer for 2 hours at 37°C. Enzymes were then heat inactivated for 20 min, at 68°C, DNA was phenol/chloroform extracted and ethanol-precipitated and its concentration and integrity were estimated by migration on a 1% agarose gel. 20ng of the BamHI and EcoRI digested pEGFP-N1
- 10 vector were added to 50ng of BamHI-EcoRI digested PG1 cDNAs. Ligation was performed over night at 13°C using 0.5units of T4 DNA ligase (BOEHRINGER ref 84333623) in a final volume of 20µl containing 1X ligase buffer. The ligation reaction mix was desalted by dialysis against water (MILLIPORE ref VSWP01300) for 30min. at room temperature. One fifth of the desalted ligation reaction was electroporated in 25µl of competent cells ElectroMAX DH10B (GIBCO BRL ref
- 15 18290-015) using a resistance of 126 Ohms, capacitance of 50µF, and voltage of 2.5KV. Bacteria were then incubated in 500µl of SOB medium for 30min at 37°C. One fifth was plated on LB AGAR containing 40µg/µl KANAMYCINE (SIGMA ref K4000) and incubated over night at 37°C. Plasmid DNA was prepared from an overnight liquid culture of individual colonies and sequenced. Among the different forms identified 3 were used :

- 20 A) PG1 which includes all the coding exons from exon 1 to 8.
- B) PG1/1-4 which corresponds to an alternative messenger which is due to an alternative splicing, joining exon 1 to exon 4, and resulting in the absence of exons 2 and 3.
- C) PG1/1-5 which corresponds to an alternative messenger which is due to an alternative splicing, joining exon 1 to exon 5, and resulting in the absence of exons 2, 3 and 4.
- 25 D) Vector PG1/1-7 : A cDNA insert encoding for a truncated protein was synthesized by PCR amplification, using primers PG15Eco3 and PG1mut29Bam (5' CCTGGATCCCCTCCATCGTCTTCCCTT 3') and vector PG1 as a template. The resulting PCR product was cloned following the same protocol as described above.

Transfection of the PG1 expression vectors in human prostate cell lines.

- 30 The DNA/lipofectin solution was prepared as followed: 1.5µl of lipofectin (GIBCO BRL ref 18292-011) was diluted in 100µl of OPTI-MEM medium (GIBCO BRL ref 31985-018), and incubated for 30min. at room temperature before being mixed to 0.5µg of vector diluted in 100µl of OPTI-MEM medium and incubated for 15 min. at room temperature. Cells were inoculated in RPMI1640 medium (Gibco BRL ref 61870-010) containing 5% fetal calf serum (Dutscher ref P30-
- 35 3302) on slides (NUNC Lab-Tek ref 177402A) and grown at 37°C in 5%CO₂. Cells reaching 40-60% confluency were rinsed with 300µl OPTI-MEM medium and incubated with the DNA/lipofectin

solution for 6 hours at 37°C. The medium containing DNA was replaced by medium supplemented in fetal calf serum and cells were incubated for at least 36 hours at 37°C. Slides were rinsed in PBS and cells were fixed in ethanol, treated with Propidium iodide, and examined with a fluorescence microscope using a double-pass filter set for FITC/PI.

5 After transfection of PG1 and PG1/1-4 in both the normal and tumoural prostatic cell line, green fluorescence was detected into and around the nucleus (Figures 10 and 11). This result shows that the PG1 protein is localised in the nucleus and/or the nuclear membrane. Furthermore, it suggests that exons 2 and 3 are dispensable for translocation of PG1 to the nucleus. In addition, no difference in the intracellular localisation of these two forms was detected between the tumoral and the normal
10 prostatic cell line.

On the contrary, transfection experiments using PG1/1-5 show that this form is cytoplasmic in the normal prostatic cell line PNT2A. It suggests that exon 4 might be important for the regulation of the translocation to the nucleus. Interestingly, similar transfection experiments in the tumoral cell line PC3 show that PG1/1-5 remains nuclear and or perinuclear (Figure 12). This result shows that there is
15 an abnormality in the regulation of the intracellular localization of the PG1 isoforms in this tumoral cell line. Furthermore, it indicates that the normal function of PG1 can be altered indirectly in prostatic tumors by an abnormality in the regulation of its intracellular location.

Finally, a non-sense mutation has been identified in the prostatic tumoural cell line LNCaP, in exon 6 of PG1 (SEQ ID NO: 69). This mutation is responsible for the production of a truncated
20 protein (SEQ ID NO: 70). To determine the intracellular location of this truncated protein, PG1/1-7 and PG1 were transfected in the normal prostatic cell line PNT2A. Comparison of the fluorescence detected in both sets of experiments clearly showed that the truncated form was localised in the cytoplasm as the non-truncated protein was located in and/or around the nucleus (Figure 13). This result indicates that this mutated PG1 is translated in a truncated protein which is unable to reach the
25 nucleus. It also suggests that exons 7 and 8 may play an important role in the regulation of the intracellular localisation of PG1. Furthermore, it supports the previous hypothesis that an altered regulation of PG1 intracellular localisation might be involved in prostate tumorigenesis.

	pEGFP N1	PG1	PG11-4	PG11-5	PG11-7
Transfection PNT2 06/17/98	NA	nuclear	nuclear	ND	ND
Transfection PNT2 06/30/98	cytoplasmic	nuclear	nuclear	ND	ND

Transfection PNT2 07/16/98	cytoplasmic	NA	NA	cytoplasmic	ND
Transfection PC3 07/16/98	NA	nuclear	nuclear	nuclear	ND
Transfection PC3 07/16/98 bis	cytoplasmic	nuclear	NA	NA	ND

Transfection PC3 08/27/98	cytoplasmic	nuclear	nuclear	nuclear	NA
Transfection PNT2 08/28/98	cytoplasmic	nuclear	NA	cytoplasmic	cytoplasmic
		All exons	X2-3 Spliced out	X2-3-4Spliced out	mut aa229

NA : Not assessable

ND : Not done

- 5 Nuclear : localized in and around the nucleus (nuclear and perinuclear localization).

Alternative Splice Species

- Alternative splicing is a common natural tool for the inhibition of function of full length gene products. Alternative splicing is known to result in enzyme isoforms, possessing different kinetic characteristics (pyruvate kinase: M1 and M2 Yamada K, Noguchi T, Biochem J. 1999 Jan1;337(Pt 1):1-11. Estrogen receptor (ER) gene is known to possess variant splicing yielding the deletions of exon 3, 5, or 7. The truncated ER protein induced from variant mRNA could mainly be exhibited as a repressor through dominant negative effects on normal ER protein (Iwase H, Omoto Y, Iwata H, Hara Y, Ando Y, Kobayashi S, Oncology 1998 Dec;55 Suppl S1:11-16) Yu et al (Yu JJ, Mu C, Dabholkar M, Guo Y, Bostick-Bruton F, Reed E, Int J Mol Med 1998 Mar;1(3):617-620) demonstrated that there is an association between alternative splicing of ERCC1, and reduction in cellular capability to repair cisplatin-DNA adduct. Munoz-Sanjuan et al (Munoz-Sanjuan I, Simandl BK, Fallon JF, Nathans J, Development 1998 Dec 14;126(Pt 2):409-421) demonstrated existence of two differentially spliced isoforms of fibroblast growth factor(FGF) type two genes that are present in non-overlapping spatial

distributions in the neural tube and adjacent structures in developing chicken embryo. One of these forms is secreted and activates the expression of HoxD13, HoxD11, Fgf-4 and BMP-2 ectopically, consistent with cFHF-2 playing a role in anterior-posterior patterning of the limb.

The CD44 is a cell adhesion molecule that is present as numerous isoforms created by mRNA
5 alternative splicing. Expression of variant isoforms of CD44 is associated with tumor growth and metastasis. (Shibuya Y, Okabayashi T, Oda K, Tanaka N, Jpn J Clin Oncol 1998 Oct;28(10):609-14) they showed that ratio of two particular isoforms is a useful indicator of prognosis in gastric and colorectal carcinoma. Zhang YF et al (Zhang YF, Jeffery S, Burchill SA, Berry PA, Kaski JC, Carter ND, Br J Cancer 1998 Nov;78(9):1141-6) showed that human endothelin receptor A is the subject to
10 alternative splicing giving at least two isoforms. The truncated receptor was expressed in all tissues and cells examined, but the level of expression varied. In melanoma cell lines and melanoma tissues, the truncated receptor gene was the major species, whereas the wild-type ETA was predominant in other tissues. Zhang et al. conclude that the function and biological significance of this truncated ETA receptor is not clear, but it may have regulatory roles for cell responses to ETs.

15

Example 8

Identification of PG1 Alternative Splice Species

The PG1 cDNA was first cloned by screening of a human prostate cDNA library. Sequence analysis of about 400 cDNA clones showed that at least 14 isoforms were present in this cDNA library. Comparison of their sequences to the genomic sequence showed that these isoforms resulted
20 from a complex set of different alternative splicing events between numerous exons (Figure 14).

To rule out the possibility of a cloning artefact generated during the cDNA library construction and to systematically identify all existing alternative splice junctions, RT-PCR experiments were performed on RNA of normal prostate as well as normal prostatic cell lines PNT1A, PNT1B and PNT2 using all the possible combinations of primers specific to the different
25 exon borders. SEQ ID NOs: 137-178. The presence of multiple PCR bands in each reaction was assessed by migration in an agarose gel. Each band was analysed by sequencing, and the presence or absence of specific splicing events, as seen in the sequence by a specific splice junction, was scored as plus or minus in Figure 15.

Furthermore, to identify aberrant splicing event in prostate tumors, similar experiments were
30 performed on RNA extracted from tumoral prostatic cell lines LnCaP (obtained from two different sources and named FCG and JMB), CaHPV, Du145 and PC3 as well as on RNA obtained from prostate tumors (ECP5 to ECP24).

As shown in the first five columns, all isoforms identified in the cDNA library were detected in RNA of normal prostate, normal prostatic cell lines or prostate tumors. In addition to the different
35 splice junctions detected in the cDNA library, 19 other splice junctions were detected in normal prostate or in normal prostatic cell lines. Two types of exon junctions (exons 3-7, exons 3b-8) were

never detected in either normal prostate, normal prostatic cell lines, prostate tumors or prostatic tumoral cell lines. Comparison between normal and tumoral samples showed the presence of 2 additional exon junctions (exons 3-8, exons 5-8) in the tumoral samples that were not detected previously in the normal samples. This result demonstrate that during tumorigenesis, the complex regulation of the PG1 splicing has been altered, resulting in an abnormal ratio of the different isoforms. It is of a specific interest since it has been shown in patients with a genetic predisposition to Wilms tumor, that an imbalance between different RNA isoforms might be involved in tumorigenesis (Bickmore et al., Science 1992, 257:325-7; Little et al, Hum Mol Genet 1995, 4:351-8).

Interestingly, comparison between normal and tumoral samples, also showed that some exon junctions are present in all normal samples, but are absent in numerous tumoral samples. It further indicates that the normal function of PG1 can be altered by an abnormality in the regulation of PG1 splicing and further support the previous hypothesis.

Furthermore, comparison between the different types of normal samples (Col.2 versus Col. 3, 4 and 5) also showed differences in the presence or absence of some exon junctions. It indicates that the transformation process necessary to the generation of these normal prostatic cell lines might result in similar alteration which further support the previous hypothesis.

Example 9

Determining the Tumor Suppressor Activity of the PG1 Gene Product, Mutants and Other PG1

Polypeptides

PG1 variants which results from either alternate splicing of the PG1 mRNA or from mutation of PG1 that introduce a stop codon (nucleotide of SEQ ID NO: 69 and protein of SEQ ID NO: 70) can no longer perform its role of tumor suppressor. It is possible and even likely that PG1 tumor suppressor role extends beyond prostate cancer to other form of malignancies. PG1 therefore represent a prime candidate for gene therapy of cancer by creating a targeting vector which knocks out the mutant and/or introduces a wild-type PG1 gene (e.g. SEQ ID NO 3 or 179) or a fragment thereof.

To validate this model, PG1 and its alternatively spliced or mutated variants are stably transfected in tumor cell line using methods described in Section VIII. The efficiency of transfection is determined by northern and western blotting; the latter is performed using antibodies prepared against PG1 synthetic peptides designed to distinguish the product of the most abundant PG1 mRNA from the alternatively spliced variants, the truncated variant, or other functional mutants. The production of synthetic peptides and of polyclonal antibodies is performed using the methods described herein in Sections III and VII. After demonstrating that PG1 and its variant are efficiently expressed in various tumor cell line preferably derived from human prostate cancer, hepatocarcinoma, lung and colon carcinoma; we the effect of this gene on the rate of cell division, DNA synthesis, ability to grow in soft agar and ability to induce tumor progression and metastasis when injected in immunologically deficient nude mice are determined.

Alternatively the PG1 gene and its variant are inserted in adenoviruses that are used to obtain a high level of expression of these genes. This method is preferred to test the effect of PG1 expression in animal that are spontaneously developing tumor. The production of specific adenoviruses is obtained using methods familiar to those with normal skills in cell and molecular biology.

II. POLYNUCLEOTIDES:

The present invention encompasses polynucleotides in the form of PG1 genomic or cDNA as well as polynucleotides for use as primers and probes in the methods of the invention. These polynucleotides may consist of, consist essentially of, or comprise a contiguous span of nucleotides of a sequence from any sequence in the Sequence Listing as well as sequences which are complementary thereto ("complements thereof"). Preferably said sequence is selected from SEQ ID NOs: 3, 112-125, 179, 182-184. The "contiguous span" is at least 6, 8, 10, 12, 15, 20, 25, 30, 50, 100, 200, or 500 nucleotides in length. It should be noted that the polynucleotides of the present invention are not limited to having the exact flanking sequences surrounding the polymorphic bases which are enumerated in Sequence Listing. Rather, it will be appreciated that the flanking sequences surrounding the biallelic markers, or any of the primers of probes of the invention which are more distant from a biallelic markers, is lengthened or shortened to any extent compatible with their intended use and the present invention specifically contemplates such sequences. It will be appreciated that the polynucleotides referred to in the Sequence Listing is of any length compatible with their intended use. Also the flanking regions outside of the contiguous span need not be homologous to native flanking sequences which actually occur in humans. The addition of any nucleotide sequence, which is compatible with the nucleotides intended use is specifically contemplated. The contiguous span may optionally include the PG1-related biallelic marker in said sequence. Optionally either allele of the biallelic markers described above in the definition of PG1-related biallelic marker is specified as being present at the PG1-related biallelic marker.

The invention also relates to polynucleotides that hybridize, under conditions of high or intermediate stringency, to a polynucleotide of a sequence from any sequence in the Sequence Listing as well as sequences, which are complementary thereto. Preferably said sequence is selected from SEQ ID NOs: 3, 112-125, 179, 182-184. Preferably such polynucleotides is at least 6, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 200, or 500 nucleotides in length. Preferred polynucleotides comprise an PG1-related biallelic marker. Optionally either allele of the biallelic markers described above in the definition of PG1-related biallelic marker is specified as being present at the biallelic marker site. Conditions of high and intermediate stringency are further described in Section X.C.4, below.

The invention embodies polynucleotides which encode an entire human, mouse or mammalian PG1 protein, or fragments thereof. Generally the polynucleotides of the invention

comprise the naturally occurring nucleotide sequence of the PG1. However, any naturally occurring silent codon variation or other silent codon variation can be employed to encode the PG1 amino acids sequence. As for those amino acids which are changed or added to the PG1 gene for any embodiment of the invention which requires the expression of a nucleotide sequence, the nucleic acid sequences
5 generally will be chosen to optimize expression in the specific human or non-human animal system in which the polynucleotide is intended to be used, making use of known codon preferences. The PG1 polynucleotides of the invention can be the native nucleotide sequence which encodes a human, mouse, or mammalian PG1 protein, preferably the PG1 polynucleotide sequence of SEQ ID NOs: 3, 112-125, 179, 182-184, and the compliments thereof. The polynucleotides of the invention include
10 those which encode PG1 polypeptides with a contiguous stretch of at least 8, 10, 12, 15, 20, 25, 30, 50, 100 or 200 amino acids from SEQ ID NOs: 4, 5, 70, 74, and 125-136, as well as any other human, mouse or mammalian PG1 polypeptide. In addition the present invention encompasses polynucleotides which comprise a contiguous stretch of at least 8, 10, 12, 15, 20, 25, 30, 50, 100, 200, 500 nucleotides of a human, mouse or mammalian PG1 genomic sequence as well as complete human,
15 mouse, or mammalian PG1 genes, preferably of SEQ ID NOs: 179, 182, 183, and the compliments thereof.

The present invention encompasses polynucleotides which consist of, consist essentially of, or comprise a contiguous stretch of at least 8, 10, 12, 15, 20, 25, 30, 50, 100, 200, or 500 nucleotides of a human, mouse or mammalian PG1 cDNA sequences as well as an entire human, mouse, or
20 mammalian PG1 cDNA. The cDNA species and polynucleotide fragments comprised by the polynucleotides of the invention include the predominant species derived from any human, mouse or mammal source, preferably SEQ ID NOs: 3, 184, and the compliments thereof. In addition, the polynucleotides of the invention comprise cDNA species, and fragments thereof, that result from the alternative splicing of PG1 transcripts in any human, mouse or other mammal, preferably the cDNA
25 species of SEQ ID NOs: 112-124, and compliments thereof. Moreover, the invention encompasses cDNA species and other polynucleotides which consist of or comprise the polynucleotides which span a splice junction, preferably including any one of SEQ ID NOs: 137 to 178, and the compliments thereof; more preferably any one of SEQ ID NOs: 137 to 149, 151 to 169, 171 to 178, and the compliments thereof. The polynucleotides of the invention also include cDNA and other
30 polynucleotides which comprise two covalently linked PG1 exons, derived from a single human, mouse or mammalian species, immediately adjacent to one another in the order shown, and selected from the following pairs of PG1 exons: 1:2, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 2:3, 2:4, 2:5, 2:6, 2:7, 2:8, 3:4, 3:5, 3:6, 3:7, 3:8, 4:5, 4:6, 4:7, 4:8, 5:6, 5:7, 5:8, 6:7, 6:8, 7:8, 1:1bis, 1bis:2, 1bis:3, 1bis:4, 1bis:5, 1bis:6, 1bis:7, 1bis:8, 3:3bis, 3bis:4, 3bis:5, 3bis:6, 3bis:7, 3bis:8, 5:5bis, 5bis:6, 5bis:7, 5bis:8,
35 1:6bis, 2:6bis, 3:6bis, 4:6bis, 5:6bis, 6bis:7, 6bis:8, and the compliments thereof. In a preferred embodiment the sequences of the PG1 exons in each of the pairs of exons is selected as follows:

exon 1 – SEQ ID NO: 100; exon 2 – SEQ ID NO: 101; exon 3 – SEQ ID NO: 102;
 exon 4 – SEQ ID NO: 103; exon 5 – SEQ ID NO: 104; exon 6 – SEQ ID NO: 105;
 exon 7 – SEQ ID NO: 106; exon 8 – SEQ ID NO: 107; exon 1bis – SEQ ID NO: 108;
 exon 3bis – SEQ ID NO: 109; exon 5bis – SEQ ID NO: 110; and

5 exon 6bis – SEQ ID NO: 111. Because of the 8 different polyadenylation sites in exon 8, any cDNA or polynucleotide of the invention comprising a human cDNA fragment encompassing exon 8 is truncated such that only the first 330 nucleotides, 699 nucleotides, 833 nucleotides, 1826 nucleotides, 2485 nucleotides, 2805 nucleotides, 4269 nucleotides or 4315 nucleotides of exon 8 shown in SEQ ID NO: 107 are present.

10 The primers of the present invention is designed from the disclosed sequences for any method known in the art. A preferred set of primers is fashioned such that the 3' end of the contiguous span of identity with the sequences of the Sequence Listing is present at the 3' end of the primer. Such a configuration allows the 3' end of the primer to hybridize to a selected nucleic acid sequence and dramatically increases the efficiency of the primer for amplification or sequencing reactions. Allele
 15 specific primers is designed such that a biallelic marker is at the 3' end of the contiguous span and the contiguous span is present at the 3' end of the primer. Such allele specific primers tend to selectively prime an amplification or sequencing reaction so long as they are used with a nucleic acid sample that contains one of the two alleles present at a biallelic marker. The 3' end of primer of the invention is located within or at least 2, 4, 6, 8, 10, 12, 15, 18, 20, 25, 50, 100, 250, 500, or 1000 nucleotides
 20 upstream of an PG1-related biallelic marker in said sequence or at any other location which is appropriate for their intended use in sequencing, amplification or the location of novel sequences or markers.

Preferred amplification primers include the polynucleotides disclosed in SEQ ID NOs: 39-56, and 63-68. Additional preferred amplification primers for particular non-genic PG1-related biallelic
 25 markers are listed as follows by the internal reference number for the marker and the SEQ ID NOs for the PU and RP amplification primers respectively:

4-14-107 use SEQ ID NOs 339 and 382; 4-14-317 use SEQ ID NOs 339 and 382;
 4-14-35 use SEQ ID NOs 339 and 382; 4-20-149 use SEQ ID NOs 340 and 383;
 4-22-174 use SEQ ID NOs 341 and 384; 4-22-176 use SEQ ID NOs 341 and 384;
 30 4-26-60 use SEQ ID NOs 342 and 385; 4-26-72 use SEQ ID NOs 342 and 385;
 4-3-130 use SEQ ID NOs 343 and 386; 4-38-63 use SEQ ID NOs 344 and 387;
 4-38-83 use SEQ ID NOs 344 and 387; 4-4-152 use SEQ ID NOs 345 and 388; 4-4-187 use
 SEQ ID NOs 345 and 388; 4-4-288 use SEQ ID NOs 345 and 388;
 4-42-304 use SEQ ID NOs 346 and 389; 4-42-401 use SEQ ID NOs 346 and 389;
 35 4-43-328 use SEQ ID NOs 347 and 390; 4-43-70 use SEQ ID NOs 347 and 390;
 4-50-209 use SEQ ID NOs 348 and 391; 4-50-293 use SEQ ID NOs 348 and 391;

- 4-50-323 use SEQ ID NOs 348 and 391; 4-50-329 use SEQ ID NOs 348 and 391;
 4-50-330 use SEQ ID NOs 348 and 391; 4-52-163 use SEQ ID NOs 349 and 392;
 4-52-88 use SEQ ID NOs 349 and 392; 4-53-258 use SEQ ID NOs 350 and 393;
 4-54-283 use SEQ ID NOs 351 and 394; 4-54-388 use SEQ ID NOs 351 and 394;
 5 4-55-70 use SEQ ID NOs 352 and 395; 4-55-95 use SEQ ID NOs 352 and 395;
 4-56-159 use SEQ ID NOs 353 and 396; 4-56-213 use SEQ ID NOs 353 and 396;
 4-58-289 use SEQ ID NOs 354 and 397; 4-58-318 use SEQ ID NOs 354 and 397;
 4-60-266 use SEQ ID NOs 355 and 398; 4-60-293 use SEQ ID NOs 355 and 398;
 4-84-241 use SEQ ID NOs 356 and 399; 4-84-262 use SEQ ID NOs 356 and 399;
 10 4-86-206 use SEQ ID NOs 357 and 400; 4-86-309 use SEQ ID NOs 357 and 400;
 4-88-349 use SEQ ID NOs 358 and 401; 4-89-87 use SEQ ID NOs 359 and 402;
 99-123-184 use SEQ ID NOs 360 and 403; 99-128-202 use SEQ ID NOs 361 and 404; 99-128-
 275 use SEQ ID NOs 361 and 404; 99-128-313 use SEQ ID NOs 361 and 404;
 99-128-60 use SEQ ID NOs 361 and 404; 99-12907-295 use SEQ ID NOs 362 and 405; 99-
 15 130-58 use SEQ ID NOs 363 and 406; 99-134-362 use SEQ ID NOs 364 and 407; 99-140-130
 use SEQ ID NOs 365 and 408; 99-1462-238 use SEQ ID NOs 366 and 409; 99-147-181 use
 SEQ ID NOs 367 and 410; 99-1474-156 use SEQ ID NOs 368 and 411; 99-1474-359 use SEQ
 ID NOs 368 and 411;
 99-1479-158 use SEQ ID NOs 369 and 412;
 20 99-1479-379 use SEQ ID NOs 369 and 412; 99-148-129 use SEQ ID NOs 370 and 413;
 99-148-132 use SEQ ID NOs 370 and 413; 99-148-139 use SEQ ID NOs 370 and 413;
 99-148-140 use SEQ ID NOs 370 and 413; 99-148-182 use SEQ ID NOs 370 and 413;
 99-148-366 use SEQ ID NOs 370 and 413; 99-148-76 use SEQ ID NOs 370 and 413; 99-1480-
 290 use SEQ ID NOs 371 and 414;
 25 99-1481-285 use SEQ ID NOs 372 and 415;
 99-1484-101 use SEQ ID NOs 373 and 416;
 99-1484-328 use SEQ ID NOs 373 and 416;
 99-1485-251 use SEQ ID NOs 374 and 417;
 99-1490-381 use SEQ ID NOs 375 and 418;
 30 99-1493-280 use SEQ ID NOs 376 and 419; 99-151-94 use SEQ ID NOs 377 and 420;
 99-211-291 use SEQ ID NOs 378 and 421; 99-213-37 use SEQ ID NOs 379 and 422;
 99-221-442 use SEQ ID NOs 380 and 423; 99-222-109 use SEQ ID NOs 381 and 424; and the
 compliments thereof.

- Primers with their 3' ends located 1 nucleotide upstream or downstream of a PG1-related
 35 biallelic marker have a special utility in microsequencing assays. Preferred microsequencing primers
 include the polynucleotides from position 1 to position 23 and from position 25 to position 47 of SEQ

ID NOs: 21-38, and as well as the compliments thereof. Additional preferred microsequencing primers for particular non-genic PG1-related biallelic markers are listed as follows by the internal reference number for the marker and the SEQ ID NOs of the two preferred microsequencing primers:

- 4-14-107 of SEQ ID NOs 425 and 502*; 4-14-317 of SEQ ID NOs 426 and 503*;
 5 4-14-35 of SEQ ID NOs 427 and 504*; 4-20-149 of SEQ ID NOs 428* and 505;
 4-20-77 of SEQ ID NOs 429 and 506; 4-22-174 of SEQ ID NOs 430* and 507;
 4-22-176 of SEQ ID NOs 431 and 508; 4-26-60 of SEQ ID NOs 432 and 509*;
 4-26-72 of SEQ ID NOs 433 and 510; 4-3-130 of SEQ ID NOs 434 and 511*;
 4-38-63 of SEQ ID NOs 435 and 512; 4-38-83 of SEQ ID NOs 436 and 513*;
 10 4-4-152 of SEQ ID NOs 437 and 514; 4-4-187 of SEQ ID NOs 438* and 515;
 4-4-288 of SEQ ID NOs 439 and 516; 4-42-304 of SEQ ID NOs 440 and 517;
 4-42-401 of SEQ ID NOs 441* and 518; 4-43-328 of SEQ ID NOs 442 and 519;
 4-43-70 of SEQ ID NOs 443* and 520; 4-50-209 of SEQ ID NOs 444* and 521;
 4-50-293 of SEQ ID NOs 445* and 522; 4-50-323 of SEQ ID NOs 446* and 523;
 15 4-50-329 of SEQ ID NOs 447* and 524; 4-50-330 of SEQ ID NOs 448 and 525;
 4-52-163 of SEQ ID NOs 449* and 526; 4-52-88 of SEQ ID NOs 450* and 527;
 4-53-258 of SEQ ID NOs 451 and 528*; 4-54-283 of SEQ ID NOs 452* and 529;
 4-54-388 of SEQ ID NOs 453 and 530; 4-55-70 of SEQ ID NOs 454 and 531*;
 4-55-95 of SEQ ID NOs 455* and 532; 4-56-159 of SEQ ID NOs 456* and 533;
 20 4-56-213 of SEQ ID NOs 457 and 534; 4-58-289 of SEQ ID NOs 458* and 535;
 4-58-318 of SEQ ID NOs 459* and 536; 4-60-266 of SEQ ID NOs 460* and 537;
 4-60-293 of SEQ ID NOs 461* and 538; 4-84-241 of SEQ ID NOs 462 and 539*;
 4-84-262 of SEQ ID NOs 463 and 540; 4-86-206 of SEQ ID NOs 464 and 541*;
 4-86-309 of SEQ ID NOs 465 and 542; 4-88-349 of SEQ ID NOs 466 and 543.;
 25 4-89-87 of SEQ ID NOs 467* and 544.; 99-123-184 of SEQ ID NOs 468 and 545;
 99-128-202 of SEQ ID NOs 469 and 546; 99-128-275 of SEQ ID NOs 470 and 547;
 99-128-313 of SEQ ID NOs 471 and 548; 99-128-60 of SEQ ID NOs 472* and 549;
 99-12907-295 of SEQ ID NOs 473 and 550*;
 99-130-58 of SEQ ID NOs 474* and 551*;
 30 99-134-362 of SEQ ID NOs 475 and 552*; 99-140-130 of SEQ ID NOs 476* and 553*;
 99-1462-238 of SEQ ID NOs 477* and 554; 99-147-181 of SEQ ID NOs 478 and 555*;
 99-1474-156 of SEQ ID NOs 479 and 556*; 99-1474-359 of SEQ ID NOs 480 and 557;
 99-1479-158 of SEQ ID NOs 481* and 558; 99-1479-379 of SEQ ID NOs 482 and 559;
 99-148-129 of SEQ ID NOs 483 and 560; 99-148-132 of SEQ ID NOs 484 and 561;
 35 99-148-139 of SEQ ID NOs 485 and 562; 99-148-140 of SEQ ID NOs 486 and 563;
 99-148-182 of SEQ ID NOs 487 and 564*; 99-148-366 of SEQ ID NOs 488 and 565;

99-148-76 of SEQ ID NOs 489 and 566; 99-1480-290 of SEQ ID NOs 490 and 567*;
 99-1481-285 of SEQ ID NOs 491 and 568*; 99-1484-101 of SEQ ID NOs 492 and 569;
 99-1484-328 of SEQ ID NOs 493* and 570;
 99-1485-251 of SEQ ID NOs 494 and 571*;
 5 99-1490-381 of SEQ ID NOs 495* and 572;
 99-1493-280 of SEQ ID NOs 496 and 573*;
 99-151-94 of SEQ ID NOs 497 and 574*; 99-211-291 of SEQ ID NOs 498* and 575;
 99-213-37 of SEQ ID NOs 499 and 576; 99-221-442 of SEQ ID 500 and 577;
 99-222-109 of SEQ ID NOs 501* and 578; and compliments thereof.

- 10 Additional preferred microsequencing primers for particular genic PG1-related biallelic markers include a polynucleotide selected from the group consisting of the nucleotide sequences from position N-X to position N-1 of SEQ ID NO:179, nucleotide sequences from position N+1 to position N+X of SEQ ID NO:179, and the compliments thereof, wherein X is equal to 15, 18, 20, 25, 30, or a range of 15 to 30, and N is equal to one of the following values: 2159; 2443; 4452; 5733; 8438;
 15 11843; 1983; 12080; 12221; 12947; 13147; 13194; 13310; 13342; 13367; 13594; 13680; 13902; 16231; 16388; 17608; 18034; 18290; 18786; 22835; 22872; 25183; 25192; 25614; 26911; 32703; 34491; 34756; 34934; 5160; 39897; 40598; 40816; 40947; 45783; 47929; 48206; 48207; 49282; 50037; 50054; 50101; 50220; 50440; 50562; 50653; 50660; 50745; 50885; 51249; 51333; 51435; 51468; 51515; 51557; 51566; 51632; 51666; 52016; 52096; 52151; 52282; 52348; 52410; 52580;
 20 52712; 52772; 52860; 53092; 53272; 53389; 53511; 53600; 53665; 53815; 54365; and 54541.

The probes of the present invention is designed from the disclosed sequences for any method known in the art, particularly methods which allow for testing if a particular sequence or marker disclosed herein is present. A preferred set of probes is designed for use in the hybridization assays of the invention in any manner known in the art such that they selectively bind to one allele of a biallelic
 25 marker, but not the other under any particular set of assay conditions. Preferred hybridization probes may consists of, consist essentially of, or comprise a contiguous span which ranges in length from 8, 10, 12, 15, 18 or 20 to 25, 35, 40, 50, 60, 70, or 80 nucleotides, or be specified as being 12, 15, 18, 20, 25, 35, 40, or 50 nucleotides in length and including a PG1-related biallelic marker of said sequence. Optionally either of the two alleles specified in the definition of PG1-related biallelic marker is
 30 specified as being present at the biallelic marker site. Optionally, said biallelic marker is within 6, 5, 4, 3, 2, or 1 nucleotides of the center of the hybridization probe or at the center of said probe. A preferred set of hybridization probes is disclosed in SEQ ID NOs: 21-38, 57-62, 185-338, and the compliments thereof. Another particularly preferred set of hybridization probes includes the polynucleotides from position X to position Y of any one of SEQ ID NOs: 21-38, 57-62, 185-338, or
 35 the compliments thereof, wherein X is equal to 5, 8, 10, 12, 14, 16, 18 or a range of 5 to 18, and Y is equal to 30, 32, 34, 36, 38, 40, 43 or a range of 30 to 43; preferably X equals 12 and Y equals 36.

Additional preferred hybridization probes for particular genic PGI-related biallelic markers include a polynucleotide selected from the group consisting of the nucleotide sequences from position N-X to position N+Y of SEQ ID NO:179, and the compliments thereof, wherein X is equal to 8, 10, 12, 15, 20, 25, or a range of 8 to 30, Y is equal to 8, 10, 12, 15, 20, 25, or a range of 8 to 30, and N is equal to
5 one of the following values: 2159; 2443; 4452; 5733; 8438; 11843; 1983; 12080; 12221; 12947; 13147; 13194; 13310; 13342; 13367; 13594; 13680; 13902; 16231; 16388; 17608; 18034; 18290; 18786; 22835; 22872; 25183; 25192; 25614; 26911; 32703; 34491; 34756; 34934; 5160; 39897; 40598; 40816; 40947; 45783; 47929; 48206; 48207; 49282; 50037; 50054; 50101; 50220; 50440; 50562; 50653; 50660; 50745; 50885; 51249; 51333; 51435; 51468; 51515; 51557; 51566; 51632;
10 51666; 52016; 52096; 52151; 52282; 52348; 52410; 52580; 52712; 52772; 52860; 53092; 53272; 53389; 53511; 53600; 53665; 53815; 54365; and 54541; wherein the nucleotide at position N is selected from one of the two alleles specified in the definition of PGI-related biallelic marker at the biallelic marker site at position N.

Any of the polynucleotides of the present invention can be labeled, if desired, by
15 incorporating a label detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive substances, fluorescent dyes or biotin. Preferably, polynucleotides are labeled at their 3' and 5' ends. A label can also be used to capture the primer, so as to facilitate the immobilization of either the primer or a primer extension product, such as amplified DNA, on a solid support. A capture label is attached to the primers or
20 probes and can be a specific binding member which forms a binding pair with the solid's phase reagent's specific binding member (e.g. biotin and streptavidin). Therefore depending upon the type of label carried by a polynucleotide or a probe, it is employed to capture or to detect the target DNA. Further, it will be understood that the polynucleotides, primers or probes provided herein, may, themselves, serve as the capture label. For example, in the case where a solid phase reagent's binding
25 member is a nucleic acid sequence, it is selected such that it binds a complementary portion of a primer or probe to thereby immobilize the primer or probe to the solid phase. In cases where a polynucleotide probe itself serves as the binding member, those skilled in the art will recognize that the probe will contain a sequence or "tail" that is not complementary to the target. In the case where a polynucleotide primer itself serves as the capture label, at least a portion of the primer will be free to
30 hybridize with a nucleic acid on a solid phase. DNA Labeling techniques are well known to the skilled technician.

Any of the polynucleotides, primers and probes of the present invention can be conveniently immobilized on a solid support. Solid supports are known to those skilled in the art and include the walls of wells of a reaction tray, test tubes, polystyrene beads, magnetic beads, nitrocellulose strips,
35 membranes, microparticles such as latex particles, sheep (or other animal) red blood cells, duracytes® and others. The solid support is not critical and can be selected by one skilled in the art. Thus, latex

particles, microparticles, magnetic or non-magnetic beads, membranes, plastic tubes, walls of microtiter wells, glass or silicon chips, sheep (or other suitable animal's) red blood cells and duracytes are all suitable examples. Suitable methods for immobilizing nucleic acids on solid phases include ionic, hydrophobic, covalent interactions and the like. A solid support, as used herein, refers to any material which is insoluble, or can be made insoluble by a subsequent reaction. The solid support can be chosen for its intrinsic ability to attract and immobilize the capture reagent. Alternatively, the solid phase can retain an additional receptor which has the ability to attract and immobilize the capture reagent. The additional receptor can include a charged substance that is oppositely charged with respect to the capture reagent itself or to a charged substance conjugated to the capture reagent. As yet another alternative, the receptor molecule can be any specific binding member which is immobilized upon (attached to) the solid support and which has the ability to immobilize the capture reagent through a specific binding reaction. The receptor molecule enables the indirect binding of the capture reagent to a solid support material before the performance of the assay or during the performance of the assay. The solid phase thus can be a plastic, derivatized plastic, magnetic or non-magnetic metal, glass or silicon surface of a test tube, microtiter well, sheet, bead, microparticle, chip, sheep (or other suitable animal's) red blood cells, duracytes® and other configurations known to those of ordinary skill in the art. The polynucleotides of the invention can be attached to or immobilized on a solid support individually or in groups of at least 2, 5, 8, 10, 12, 15, 20, or 25 distinct polynucleotides of the inventions to a single solid support. In addition, polynucleotides other than those of the invention may attached to the same solid support as one or more polynucleotides of the invention.

Any polynucleotide provided herein is attached in overlapping areas or at random locations on the solid support. Alternatively the polynucleotides of the invention is attached in an ordered array wherein each polynucleotide is attached to a distinct region of the solid support which does not overlap with the attachment site of any other polynucleotide. Preferably, such an ordered array of polynucleotides is designed to be "addressable" where the distinct locations are recorded and can be accessed as part of an assay procedure. Addressable polynucleotide arrays typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different known locations. The knowledge of the precise location of each polynucleotides location makes these "addressable" arrays particularly useful in hybridization assays. Any addressable array technology known in the art can be employed with the polynucleotides of the invention. One particular embodiment of these polynucleotide arrays is known as the Genechips™, and has been generally described in US Patent 5,143,854; PCT publications WO 90/15070 and 92/10092. These arrays may generally be produced using mechanical synthesis methods or light directed synthesis methods, which incorporate a combination of photolithographic methods and solid phase oligonucleotide synthesis (Fodor et al., Science, 251:767-777, 1991). The immobilization of arrays of oligonucleotides on solid

supports has been rendered possible by the development of a technology generally identified as "Very Large Scale Immobilized Polymer Synthesis" (VLSIPS™) in which, typically, probes are immobilized in a high density array on a solid surface of a chip. Examples of VLSIPS™ technologies are provided in US Patents 5,143,854 and 5,412,087 and in PCT Publications WO 90/15070, WO 5 92/10092 and WO 95/11995, which describe methods for forming oligonucleotide arrays through techniques such as light-directed synthesis techniques. In designing strategies aimed at providing arrays of nucleotides immobilized on solid supports, further presentation strategies were developed to order and display the oligonucleotide arrays on the chips in an attempt to maximize hybridization patterns and sequence information. Examples of such presentation strategies are disclosed in PCT 10 Publications WO 94/12305, WO 94/11530, WO 97/29212 and WO 97/31256.

Oligonucleotide arrays may comprise at least one of the sequences selected from the group consisting of SEQ ID NOs: 3, 21-38, 57-62, 100-124, 179, 185-338, the preferred hybridization probes for genic PG1-related biallelic markers described above; and the sequences complementary thereto; or a fragment thereof of at least 15 consecutive nucleotides for determining whether a sample 15 contains one or more alleles of the biallelic markers of the present invention. Oligonucleotide arrays may also comprise at least one of the sequences selected from the group consisting of SEQ ID NOs: 179, 339-424; and the sequences complementary thereto or a fragment thereof of at least 15 consecutive nucleotides for amplifying one or more alleles of the PG1-related biallelic markers. In other embodiments, arrays may also comprise at least one of the sequences selected from the group 20 consisting of SEQ ID 425-578, the preferred microsequencing primers for genic PG1-related biallelic markers described above; and the sequences complementary thereto or a fragment thereof of at least 15 consecutive nucleotides for conducting microsequencing analyses to determine whether a sample contains one or more alleles of PG1-related biallelic marker.

The present invention further encompasses polynucleotide sequences that hybridize to any 25 one of SEQ ID NOs: 3, 69, 100-112, or 179-184 under conditions of high or intermediate stringency as described below:

(i) By way of example and not limitation, procedures using conditions of high stringency are as follows: Prehybridization of filters containing DNA is carried out for 8 h to overnight at 65°C in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 30 0.02% BSA, and 500 µg/ml denatured salmon sperm DNA. Filters are hybridized for 48 h at 65°C, the preferred hybridization temperature, in prehybridization mixture containing 100 µg/ml denatured salmon sperm DNA and 5-20 X 10⁶ cpm of ³²P-labeled probe. Alternatively, the hybridization step can be performed at 65°C in the presence of SSC buffer, 1 x SSC corresponding to 0.15M NaCl and 0.05 M Na citrate. Subsequently, filter washes can be done at 37°C for 1 h in a solution containing 35 2X SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA, followed by a wash in 0.1X SSC at 50°C for 45 min. Alternatively, filter washes can be performed in a solution containing 2 x SSC and 0.1% SDS,

or 0.5 x SSC and 0.1% SDS, or 0.1 x SSC and 0.1% SDS at 68°C for 15 minute intervals. Following the wash steps, the hybridized probes are detectable by autoradiography. Other conditions of high stringency which is used are well known in the art and as cited in Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Second Edition, Cold Spring Harbor Press, N.Y., pp. 9.47-9.57; and
5 Ausubel et al., 1989, Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y. Preferably, such sequences encode a homolog of a polypeptide encoded by one of ORF2 to ORF1297. In one embodiment, such sequences encode a mammalian PG1 polypeptide.

(ii) By way of example and not limitation, procedures using conditions of intermediate stringency are as follows: Filters containing DNA are prehybridized, and then hybridized at a
10 temperature of 60°C in the presence of a 5 x SSC buffer and labeled probe. Subsequently, filters washes are performed in a solution containing 2x SSC at 50°C and the hybridized probes are detectable by autoradiography. Other conditions of intermediate stringency which is used are well known in the art and as cited in Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Second Edition, Cold Spring Harbor Press, N.Y., pp. 9.47-9.57; and Ausubel et al., 1989, Current
15 Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y. Preferably, such sequences encode a homolog of a polypeptide encoded by one of SEQ ID NOs: 3, 69, 100-112, or 179-184. In one embodiment, such sequences encode a mammalian PG1 polypeptide.

The present invention also encompasses diagnostic kits comprising one or more polynucleotides of the invention with a portion or all of the necessary reagents and instructions for
20 genotyping a test subject by determining the identity of a nucleotide at a PG1-related biallelic marker. The polynucleotides of a kit may optionally be attached to a solid support, or be part of an array or addressable array of polynucleotides. The kit may provide for the determination of the identity of the nucleotide at a marker position by any method known in the art including, but not limited to, a sequencing assay method, a microsequencing assay method, a hybridization assay method, or an allele
25 specific amplification method. Optionally such a kit may include instructions for scoring the results of the determination with respect to the test subjects' risk of contracting a cancer or prostate cancer, or likely response to an anti-cancer agent or anti-prostate cancer agent, or chances of suffering from side effects to an anti-cancer agent or anti-prostate cancer agent.

Use of PG1 Nucleic Acids as Reagents

30 The PG1 genomic DNA of SEQ ID NO: 179, the PG1 cDNA of SEQ ID NO: 3, 112-124 and PG1 alleles responsible for a detectable phenotype (such as those obtainable by the methods of Example 12, and SEQ ID NO:69) can be used to prepare PCR primers for use in diagnostic techniques or genetic engineering methods such as those described above. Example 10 describes the use of the PG1 genomic DNA of SEQ ID NO: 179, the PG1 cDNA of SEQ ID NO: 3, 112-124 and PG1 alleles responsible for a
35 detectable phenotype (such as those obtainable by the methods of Example 12) in PCR amplification procedures.

Example 10

The PG1 genomic DNA of SEQ ID NO: 179, the PG1 cDNA of SEQ ID NO: 3, and PG1 alleles responsible for a detectable phenotype (such as those obtainable by the methods of Example 12) is used to prepare PCR primers for a variety of applications, including isolation procedures for cloning nucleic acids capable of hybridizing to such sequences, diagnostic techniques and forensic techniques. The PCR primers comprise at least 10 consecutive bases of the PG1 genomic DNA of SEQ ID NO: 179, the PG1 cDNA of SEQ ID NO: 3, 112-124 and PG1 alleles responsible for a detectable phenotype (such as those obtainable by the methods of Example 12) or the sequences complementary thereto. Preferably, the PCR primers comprise at least 12, 15, or 17 consecutive bases of these sequences. More preferably, the PCR primers comprise at least 20-30 consecutive bases of the PG1 genomic DNA of SEQ ID NO: 179, the PG1 cDNA of SEQ ID NO: 3, 112-124 and PG1 alleles responsible for a detectable phenotype (such as those obtainable by the methods of Example 12) or the sequences complementary thereto. In some embodiments, the PCR primers may comprise more than 30 consecutive bases of the PG1 genomic DNA of SEQ ID NO: 179, the PG1 cDNA of SEQ ID NO: 3, 112-124 and PG1 alleles responsible for a detectable phenotype (such as those obtainable by the methods of Example 12) or the sequences complementary thereto. It is preferred that the primer pairs to be used together in a PCR amplification have approximately the same G/C ratio, so that melting temperatures are approximately the same.

A variety of PCR techniques are familiar to those skilled in the art. For a review of PCR technology, see Molecular Cloning to Genetic Engineering White, B.A. Ed. in Methods in Molecular Biology 67: Humana Press, Totowa 1997. In each of these PCR procedures, PCR primers on either side of the nucleic acid sequences to be amplified are added to a suitably prepared nucleic acid sample along with dNTPs and a thermostable polymerase such as Taq polymerase, Pfu polymerase, or Vent polymerase. The nucleic acid in the sample is denatured and the PCR primers are specifically hybridized to complementary nucleic acid sequences in the sample. The hybridized primers are extended. Thereafter, another cycle of denaturation, hybridization, and extension is initiated. The cycles are repeated multiple times to produce an amplified fragment containing the nucleic acid sequence between the primer sites.

The polynucleotides of the Invention also encompass vectors and DNA constructs as well as other forms of primers and probes. For a thorough description of these embodiments please see Sections VIII, X, and XI below.

III. POLYPEPTIDES

PG1 Proteins and Polypeptide Fragments

The term "PG1 polypeptides" is used herein to embrace all of the proteins and polypeptides of the present invention. Also forming part of the invention are polypeptides encoded by the polynucleotides of the invention, as well as fusion polypeptides comprising such polypeptides. The invention embodies PG1 proteins from human (SEQ ID NOs: 4, and 5), and mouse (SEQ ID NO: 74).

However, PG1 species from other varieties of mammals are expressly contemplated and is isolated using the antibodies of the present invention in conjunction with standard affinity chromatography methods as well as being expressed from the PG1 genes isolated from other mammalian sources using human and mouse PG1 nucleic acid sequences as primers and probes as well as the methods described
5 herein.

The invention also embodies PG1 proteins translated from less common alternative splice species, including SEQ ID NOs: 125-136, and PG1 proteins which result from naturally occurring mutant, particularly functional mutants of PG1, including SEQ ID NO: 70, which is identified and obtained by the described herein. The present invention also embodies polypeptides comprising a
10 contiguous stretch of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 50, or 100 amino acids of a PG1 protein. In a preferred embodiment the contiguous stretch of amino acids comprises the site of a mutation or functional mutation, including a deletion, addition, swap or truncation of the amino acids in the PG1 protein sequence. For instance, polypeptides that contain either the Arg and His residues at amino acid position 184, and polypeptides
15 that contain either the Arg or Ile residue at amino acid position 293 of the SEQ ID NO: 4 in said contiguous stretch are particularly preferred embodiments of the invention and useful in the manufacture of antibodies to detect the presence and absence of these mutations. Similarly, polypeptides with a carboxy terminus at position 228 is a particularly preferred embodiment of the invention and useful in the manufacture of antibodies to detect the presence and absence of the
20 mutation shown in SEQ ID NOs: 69 and 70.

Similarly, polypeptides that contain an peptide sequences of 8, 10, 12, 15, or 25 amino acids encoded over a naturally-occurring splice junction (the point at which two human PG1 exon (SEQ ID NOs: 100-111) are covalently linked) in said contiguous stretch are particularly preferred embodiments and useful in the manufacture of antibodies to detect the presence, localization, and
25 quantity of the various protein products of the PG1 alternative splice species.

PG1 proteins are preferably isolated from human, mouse or mammalian tissue samples or expressed from human, mouse or mammalian genes.

The PG1 polypeptides of the invention can be made using routine expression methods known in the art, see, for instance, Example 11, below. The polynucleotide encoding the desired
30 polypeptide, is ligated into an expression vector suitable for any convenient host. Both eukaryotic and prokaryotic host systems is used in forming recombinant polypeptides, and a summary of some of the more common systems are included in Sections II and VIII. The polypeptide is then isolated from lysed cells or from the culture medium and purified to the extent needed for its intended use. Purification is by any technique known in the art, for example, differential extraction, salt
35 fractionation, chromatography, centrifugation, and the like. See, for example, *Methods in Enzymology* for a variety of methods for purifying proteins.

In addition, shorter protein fragments is produced by chemical synthesis. Alternatively the proteins of the invention is extracted from cells or tissues of humans or non-human animals. Methods for purifying proteins are known in the art, and include the use of detergents or chaotropic agents to disrupt particles followed by differential extraction and separation of the polypeptides by ion exchange chromatography, affinity chromatography, sedimentation according to density, and gel electrophoresis.

Expression of the PG1 Protein

Any PG1 cDNA, including SEQ ID NO: 3, 69, 112-124, or 184 or synthetic DNAs is use as described in Example 11 below to express PG1 proteins and polypeptides.

10

Example 11

The nucleic acid encoding the PG1 protein or polypeptide to be expressed is operably linked to a promoter in an expression vector using conventional cloning technology. The PG1 insert in the expression vector may comprise the full coding sequence for the PG1 protein or a portion thereof. For example, the PG1 derived insert may encode a polypeptide comprising at least 10 consecutive amino acids of the PG1 proteins of SEQ ID NO: 4.

The expression vector is any of the mammalian, yeast, insect or bacterial expression systems known in the art, see for example Section VIII. Commercially available vectors and expression systems are available from a variety of suppliers including Genetics Institute (Cambridge, MA), Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence is optimized for the particular expression organism in which the expression vector is introduced, as explained by Hatfield, et al., U.S. Patent No. 5,082,767.

The following is provided as one exemplary method to express the PG1 protein or a portion thereof. In one embodiment, the entire coding sequence of the PG1 cDNA through the poly A signal of the cDNA are operably linked to a promoter in the expression vector. Alternatively, if the nucleic acid encoding a portion of the PG1 protein lacks a methionine to serve as the initiation site, an initiating methionine can be introduced next to the first codon of the nucleic acid using conventional techniques. Similarly, if the insert from the PG1 cDNA lacks a poly A signal, this sequence can be added to the construct by, for example, splicing out the Poly A signal from pSG5 (Stratagene) using BglI and Sall restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene). pXT1 contains the LTRs and a portion of the gag gene from Moloney Murine Leukemia Virus. The position of the LTRs in the construct allow efficient stable transfection. The vector includes the Herpes Simplex Thymidine Kinase promoter and the selectable neomycin gene. The nucleic acid encoding the PG1 protein or a portion thereof is obtained by PCR from a bacterial vector containing the PG1 cDNA of SEQ ID NO: 3 using oligonucleotide primers complementary to the PG1 cDNA or portion thereof and containing restriction endonuclease sequences for Pst I incorporated into the 5' primer and

BglII at the 5' end of the corresponding cDNA 3' primer, taking care to ensure that the sequence encoding the PG1 protein or a portion thereof is positioned properly with respect to the poly A signal. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with Bgl II, purified and ligated to pXT1, now containing a poly A signal and
5 digested with BglII.

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600ug/ml G418 (Sigma, St. Louis, Missouri).

10 Alternatively, the nucleic acids encoding the PG1 protein or a portion thereof is cloned into pED6dpc2 (Genetics Institute, Cambridge, MA). The resulting pED6dpc2 constructs is transfected into a suitable host cell, such as COS 1 cells. Methotrexate resistant cells are selected and expanded.

The above procedures may also be used to express a mutant PG1 protein responsible for a detectable phenotype or a portion thereof.

15 The expressed proteins is purified using conventional purification techniques such as ammonium sulfate precipitation or chromatographic separation based on size or charge. The protein encoded by the nucleic acid insert may also be purified using standard immunochromatography techniques. In such procedures, a solution containing the expressed PG1 protein or portion thereof, such as a cell extract, is applied to a column having antibodies against the PG1 protein or portion thereof is attached to the
20 chromatography matrix. The expressed protein is allowed to bind the immunochromatography column. Thereafter, the column is washed to remove non-specifically bound proteins. The specifically bound expressed protein is then released from the column and recovered using standard techniques.

To confirm expression of the PG1 protein or a portion thereof, the proteins expressed from host cells containing an expression vector containing an insert encoding the PG1 protein or a portion thereof
25 can be compared to the proteins expressed in host cells containing the expression vector without an insert. The presence of a band in samples from cells containing the expression vector with an insert which is absent in samples from cells containing the expression vector without an insert indicates that the PG1 protein or a portion thereof is being expressed. Generally, the band will have the mobility expected for the PG1 protein or portion thereof. However, the band may have a mobility different than
30 that expected as a result of modifications such as glycosylation, ubiquitination, or enzymatic cleavage.

Antibodies capable of specifically recognizing the expressed PG1 protein or a portion thereof is generated as described below in Section VII.

If antibody production is not possible, the nucleic acids encoding the PG1 protein or a portion thereof is incorporated into expression vectors designed for use in purification schemes employing
35 chimeric polypeptides. In such strategies the nucleic acid encoding the PG1 protein or a portion thereof is inserted in frame with the gene encoding the other half of the chimera. The other half of the chimera

is β -globin or a nickel binding polypeptide encoding sequence. A chromatography matrix having antibody to β -globin or nickel attached thereto is then used to purify the chimeric protein. Protease cleavage sites is engineered between the β -globin gene or the nickel binding polypeptide and the PG1 protein or portion thereof. Thus, the two polypeptides of the chimera is separated from one another by protease digestion.

One useful expression vector for generating β -globin chimerics is pSG5 (Stratagene), which encodes rabbit β -globin. Intron II of the rabbit β -globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis et al., (Basic Methods in Molecular Biology, L.G. Davis, M.D. Dibner, and J.F. Battey, ed., Elsevier Press, NY, 1986) and many of the methods are available from Stratagene, Life Technologies, Inc., or Promega. Polypeptide may additionally be produced from the construct using in vitro translation systems such as the In vitro ExpressTM Translation Kit (Stratagene).

IV. IDENTIFICATION OF MUTATIONS IN THE PG1 GENE WHICH ARE ASSOCIATED WITH A DETECTABLE PHENOTYPE

Mutations in the PG1 gene which are responsible for a detectable phenotype is identified by comparing the sequences of the PG1 genes from affected and unaffected individuals as described in Example 12, below. The detectable phenotype may comprise a variety of manifestations of altered PG1 function, including prostate cancer, hepatocellular carcinoma, colorectal cancer, non-small cell lung cancer, squamous cell carcinoma, or other conditions. The mutations may comprise point mutations, deletions, or insertions of the PG1 gene. The mutations may lie within the coding sequence for the PG1 protein or within regulatory regions in the PG1 gene.

Example 12

Oligonucleotide primers are designed to amplify the sequences of each of the exons or the promoter region of the PG1 gene. The oligonucleotide primers may comprise at least 10 consecutive nucleotides of the PG1 genomic DNA of SEQ ID NO:179 or the PG1 cDNA of SEQ ID NO: 3 or the sequences complementary thereto. Preferably, the oligonucleotides comprise at least 15 consecutive nucleotides of the PG1 genomic DNA of SEQ ID NO:179 or the PG1 cDNA of SEQ ID NO: 3 or the sequences complementary thereto. In some embodiments, the oligonucleotides may comprise at least 20 consecutive nucleotides of the PG1 genomic DNA of SEQ ID NO: 179 or the PG1 cDNA of SEQ ID NO:3 or the sequences complementary thereto. In other embodiments, the oligonucleotides may comprise 25 or more consecutive nucleotides of the PG1 genomic DNA of SEQ ID NO: 179 or the PG1 cDNA of SEQ ID NO: 3 or the sequences complementary thereto.

Each primer pair is used to amplify the exon or promoter region from which it is derived. Amplification is carried out on genomic DNA samples from affected patients and unaffected controls using the PCR conditions described above. Amplification products from the genomic PCRs are

subjected to automated dideoxy terminator sequencing reactions and electrophoresed on ABI 377 sequencers. Following gel image analysis and DNA sequence extraction, ABI sequence data are automatically analyzed to detect the presence of sequence variations among affected and unaffected individuals. Sequences are verified by determining the sequences of both DNA strands for each individual. Preferably, these candidate mutations are detected by comparing individuals homozygous for haplotype 5 of Figure 4 and controls not carrying haplotype 5 or related haplotypes.

Candidate polymorphisms suspected of being responsible for the detectable phenotype, such as prostate cancer or other conditions, are then verified by screening a larger population of affected and unaffected individuals using the microsequencing technique described above. Polymorphisms which exhibit a statistically significant correlation with the detectable phenotype are deemed responsible for the detectable phenotype.

Other techniques may also be used to detect polymorphisms associated with a detectable phenotype such as prostate cancer or other conditions. For example, polymorphisms is detected using single stranded conformation analyses such as those described in Orita et al., Proc. Natl. Acad. Sci. U.S.A. 86: 2776-2770 (1989). In this approach, polymorphisms are detected through altered migration on SSCA gels.

Alternatively, polymorphisms is identified using clamped denaturing gel electrophoresis, heteroduplex analysis, chemical mismatch cleavage, and other conventional techniques as described in Sheffield, V.C. et al, Proc. Natl. Acad. Sci. U.S.A 49:699-706 (1991); White, M.B. et al., Genomics 12:301-306 (1992); Grompe, M. et al., Proc. Natl. Acad. Sci. U.S.A 86:5855-5892 (1989); and Grompe, M. Nature Genetics 5:111-117 (1993).

The PG1 genes from individuals carrying PG1 mutations responsible for the detectable phenotype, or cDNAs derived therefrom, is cloned as follows. Nucleic acid samples are obtained from individuals having a PG1 mutation associated with the detectable phenotype. The nucleic acid samples are contacted with a probe derived from the PG1 genomic DNA of SEQ ID NO: 179 or the PG1 cDNA of SEQ ID NO:3. Nucleic acids containing the mutant PG1 allele are identified using conventional techniques. For example, the mutant PG1 gene, or a cDNA derived therefrom, is obtained by conducting an amplification reaction using primers derived from the PG1 genomic DNA of SEQ ID NO: 179 or the PG1 cDNA of SEQ ID NO:3. Alternatively, the mutant PG1 gene, or a cDNA derived therefrom, is identified by hybridizing a genomic library or a cDNA library obtained from an individual having a mutant PG1 gene with a detectable probe derived from the PG1 genomic DNA of SEQ ID NO: 179 or the PG1 cDNA of SEQ ID NO: 3. Alternatively, the mutant PG1 allele is obtained by contacting an expression library from an individual carrying a PG1 mutation with a detectable antibody against the PG1 proteins of SEQ ID NO: 4 or SEQ ID NO: 5 which has been prepared as described below. Those skilled in the art will appreciate that the PG1 genomic DNA of

SEQ ID NO: 179, the PG1 cDNA of SEQ ID NO: 3 and the PG1 proteins of SEQ ID NOs: 4 and 5 is used in a variety of other conventional techniques to obtain the mutant PG1 gene.

5 In another embodiment the mutant PG1 allele which causes a detectable phenotype can be isolated by obtaining a nucleic acid sample such as a genomic library or a cDNA library from an individual expressing the detectable phenotype. The nucleic acid sample can be contacted with one or more probes lying in the 8p23 region of the human genome. Nucleic acids in the sample which contain the PG1 gene can be identified by conducting sequencing reactions on the nucleic acids which hybridize to the markers from the 8p23 region of the human genome.

10 The region of the PG1 gene containing the mutation responsible for the detectable phenotype may also be used in diagnostic techniques such as those described below. For example, oligonucleotides containing the mutation responsible for the detectable phenotype is used in amplification or hybridization based diagnostics, such as those described herein, for detecting individuals suffering from the detectable phenotype or individuals at risk of developing the detectable phenotype at a subsequent time. In addition, the PG1 allele responsible for the detectable phenotype
15 is used in gene therapy as described herein. The PG1 allele responsible for the detectable phenotype may also be cloned into an expression vector to express the mutant PG1 protein as described herein.

During the search for biallelic markers associated with prostate cancer, a number of polymorphic bases were discovered which lie within the PG1 gene. The identities and positions of these polymorphic bases are listed as features in the accompanying Sequence Listing for the PG1
20 genomic DNA of SEQ ID NO: 179. The polymorphic bases is used in the above-described diagnostic techniques for determining whether an individual is at risk for developing prostate cancer at a subsequent date or suffers from prostate cancer as a result of a PG1 mutation. The identities of the nucleotides present at the polymorphic positions in a nucleic acid sample is determined using the techniques, such as microsequencing analysis, which are described above.

25 It is possible that one or more of these polymorphisms (or other polymorphic bases) is mutations which are associated with prostate cancer. To determine whether a polymorphism is responsible for prostate cancer, the frequency of each of the alleles in individuals suffering from prostate cancer and unaffected individuals is measured as described in the haplotype analysis above. Those mutations which occur at a statistically significant frequency in the affected population are
30 deemed to be responsible for prostate cancer.

cDNAs containing the identified mutant PG1 gene is prepared as described above and cloned into expression vectors as described below. The proteins expressed from the expression vectors is used to generate antibodies specific for the mutant PG1 proteins as described below. In addition, allele specific probes containing the PG1 mutation responsible for prostate cancer is used in the
35 diagnostic techniques described below.

Genes sharing homology to the PG1 gene is identified as follows.

Example 13

Alternatively, a cDNA library or genomic DNA library to be screened for genes sharing homology to the PG1 gene is obtained from a commercial source or made using techniques familiar to those skilled in the art. The cDNA library or genomic DNA library is hybridized to a detectable probe
5 comprising at least 10 consecutive nucleotides from the PG1 cDNA of SEQ ID NO:3, the PG1 genomic DNA of SEQ ID NO: 179, or the sequences complementary thereto, using conventional techniques. Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the PG1 cDNA of SEQ ID NO:3, the PG1 genomic DNA of SEQ ID NO: 179, or the sequences complementary thereto. More preferably, the probe comprises at least 20-30 consecutive nucleotides from the PG1 cDNA of
10 SEQ ID NO:3, the PG1 genomic DNA of SEQ ID NO: 179, or the sequences complementary thereto. In some embodiments, the probe comprises more than 30 nucleotides from the PG1 cDNA of SEQ ID NO:3, the PG1 genomic DNA of SEQ ID NO: 179, or the sequences complementary thereto.

Techniques for identifying cDNA clones in a cDNA library which hybridize to a given probe sequence are disclosed in Sambrook et al., Molecular Cloning: A Laboratory Manual 2d Ed., Cold
15 Spring Harbor Laboratory Press, 1989. The same techniques is used to isolate genomic DNAs sharing homology with the PG1 gene.

Briefly, cDNA or genomic DNA clones which hybridize to the detectable probe are identified and isolated for further manipulation as follows. A probe comprising at least 10 consecutive nucleotides from the PG1 cDNA of SEQ ID NO:3, the PG1 genomic DNA of SEQ ID NO: 179, or the sequences
20 complementary thereto, is labeled with a detectable label such as a radioisotope or a fluorescent molecule. Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the PG1 cDNA of SEQ ID NO:3, the PG1 genomic DNA of SEQ ID NO: 179, or the sequences complementary thereto. More preferably, the probe comprises 20-30 consecutive nucleotides from the PG1 cDNA of SEQ ID NO:3, the PG1 genomic DNA of SEQ ID NO: 179, or the sequences complementary thereto. In
25 some embodiments, the probe comprises more than 30 nucleotides from the PG1 cDNA of SEQ ID NO:3, the PG1 genomic DNA of SEQ ID NO: 179, or the sequences complementary thereto.

Techniques for labeling the probe are well known and include phosphorylation with polynucleotide kinase, nick translation, in vitro transcription, and non-radioactive techniques. The cDNAs or genomic DNAs in the library are transferred to a nitrocellulose or nylon filter and denatured.
30 After incubation of the filter with a blocking solution, the filter is contacted with the labeled probe and incubated for a sufficient amount of time for the probe to hybridize to cDNAs or genomic DNAs containing a sequence capable of hybridizing to the probe.

By varying the stringency of the hybridization conditions used to identify cDNAs or genomic DNAs which hybridize to the detectable probe, cDNAs or genomic DNAs having different levels of
35 homology to the probe can be identified and isolated. To identify cDNAs or genomic DNAs having a

high degree of homology to the probe sequence, the melting temperature of the probe is calculated using the following formulas:

For probes between 14 and 70 nucleotides in length the melting temperature T_m is calculated using the formula: $T_m = 81.5 + 16.6(\log(Na^+)) + 0.41(\text{fraction G+C}) - (600/N)$ where N is the length of the probe.

If the hybridization is carried out in a solution containing formamide, the melting temperature is calculated using the equation $T_m = 81.5 + 16.6(\log(Na^+)) + 0.41(\text{fraction G+C}) - (0.63\% \text{ formamide}) - (600/N)$ where N is the length of the probe.

Prehybridization is carried out in 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 100 μ g denatured fragmented salmon sperm DNA or 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 100 μ g denatured fragmented salmon sperm DNA, 50% formamide. The formulas for SSC and Denhardt's solutions are listed in Sambrook et al., supra.

Hybridization is conducted by adding the detectable probe to the prehybridization solutions listed above. Where the probe comprises double stranded DNA, it is denatured before addition to the hybridization solution. The filter is contacted with the hybridization solution for a sufficient period of time to allow the probe to hybridize to cDNAs or genomic DNAs containing sequences complementary thereto or homologous thereto. For probes over 200 nucleotides in length, the hybridization is carried out at 15-25 $^{\circ}$ C below the T_m . For shorter probes, such as oligonucleotide probes, the hybridization is conducted at 15-25 $^{\circ}$ C below the T_m . Preferably, for hybridizations in 6X SSC, the hybridization is conducted at approximately 68 $^{\circ}$ C. Preferably, for hybridizations in 50% formamide containing solutions, the hybridization is conducted at approximately 42 $^{\circ}$ C.

All of the foregoing hybridizations would be considered to be under "stringent" conditions.

Following hybridization, the filter is washed in 2X SSC, 0.1% SDS at room temperature for 15 minutes. The filter is then washed with 0.1X SSC, 0.5% SDS at room temperature for 30 minutes to 1 hour. Thereafter, the solution is washed at the hybridization temperature in 0.1X SSC, 0.5% SDS. A final wash is conducted in 0.1X SSC at room temperature.

cDNAs or genomic DNAs homologous to the PG1 gene which have hybridized to the probe are identified by autoradiography or other conventional techniques.

The above procedure is modified to identify cDNAs or genomic DNAs having decreasing levels of homology to the probe sequence. For example, to obtain cDNAs or genomic DNAs of decreasing homology to the detectable probe, less stringent conditions is used. For example, the hybridization temperature is decreased in increments of 5 $^{\circ}$ C from 68 $^{\circ}$ C to 42 $^{\circ}$ C in a hybridization buffer having a Na^+ concentration of approximately 1M. Following hybridization, the filter is washed with 2X SSC, 0.5% SDS at the temperature of hybridization. These conditions are considered to be "moderate" conditions above 50 $^{\circ}$ C and "low" conditions below 50 $^{\circ}$ C.

Alternatively, the hybridization is carried out in buffers, such as 6X SSC, containing formamide at a temperature of 42° C. In this case, the concentration of formamide in the hybridization buffer is reduced in 5% increments from 50% to 0% to identify clones having decreasing levels of homology to the probe. Following hybridization, the filter is washed with 6X SSC, 0.5% SDS at 50° C. These
5 conditions are considered to be "moderate" conditions above 25% formamide and "low" conditions below 25% formamide.

cDNAs or genomic DNAs which have hybridized to the probe are identified by autoradiography.

If it is desired to obtain nucleic acids homologous to the PG1 gene, such as allelic variants thereof or nucleic acids encoding proteins related to the PG1 protein, the level of homology between the
10 hybridized nucleic acid and the PG1 gene may readily be determined. To determine the level of homology between the hybridized nucleic acid and the PG1 gene, the nucleotide sequences of the hybridized nucleic acid and the PG1 gene are compared. For example, using the above methods, nucleic acids having at least 95% nucleic acid homology to the PG1 gene is obtained and identified. Similarly, by using progressively less stringent hybridization conditions one can obtain and identify nucleic acids
15 having at least 90%, at least 85%, at least 80% or at least 75% homology to the PG1 gene.

To determine whether a clone encodes a protein having a given amount of homology to the PG1 protein, the amino acid sequence of the PG1 protein is compared to the amino acid sequence encoded by the hybridizing nucleic acid. Homology is determined to exist when an amino acid sequence in the PG1 protein is closely related to an amino acid sequence in the hybridizing nucleic acid. A sequence is
20 closely related when it is identical to that of the PG1 sequence or when it contains one or more amino acid substitutions therein in which amino acids having similar characteristics have been substituted for one another. Using the above methods, one can obtain nucleic acids encoding proteins having at least 95%, at least 90%, at least 85%, at least 80% or at least 75% homology to the proteins encoded by the PG1 probe.

25 Isolation and Use of Mutant or Low Frequency PG1 Alleles from Mammalian Prostate Tumor Tissues and Cell lines

A single mutant PG1 gene was isolated from a human prostate cancer cell line. The nucleic acid sequence and amino acid sequence of this mutant PG1 are disclosed in SEQ IN NOs: 69 and 70, respectively. This mutant was found to contain a stop codon at codon position number 229, and
30 therefore results in a truncated gene product of only 228 amino acids. The present invention encompasses purified or isolated nucleic acids comprising at least 8, 10, 12, 15, 20, or 25 consecutive nucleotides of SEQ ID NO: 69, preferably containing the mutation in codon number 229. A preferred embodiment of the present invention encompasses purified or isolated nucleic acids comprising at least 8, 10, 12, 15, 20, or 25 consecutive nucleotides of SEQ ID NO: 71.

35 The present invention is also directed to methods of determining whether an individual is at risk of developing prostate cancer at a later date or whether said individual suffers from prostate

cancer as a result of a mutation in the PG1 gene comprising: obtaining a nucleic acid sample from said individual; and determining whether the nucleotides present at one or more of the polymorphic bases in the sequences selected from the group consisting of SEQ ID NOs: 69 and 71 are indicative of a risk of developing prostate cancer at a later date or indicative of prostate cancer resulting from a mutation in the PG1 gene. The present invention also includes purified or isolated nucleic acids encoding at least 4, 8, 10, 12, 15, or 20 consecutive amino acids of the polypeptide of SEQ ID NO: 70, preferably including the carboxy terminus of said polypeptide. The isolated or purified polypeptides of the invention include polypeptides comprising at least 4, 8, 10, 12, 15, or 20 consecutive amino acids of the polypeptide of SEQ ID NO: 70, preferably including the carboxy terminus of said polypeptide.

V. DIAGNOSIS OF INDIVIDUALS AT RISK FOR DEVELOPING PROSTATE CANCER OR INDIVIDUALS SUFFERING FROM PROSTATE CANCER AS A RESULT OF A MUTATION IN THE PG1 GENE

Individuals may then be screened for the presence of polymorphisms in the PG1 gene or protein which are associated with a detectable phenotype such as cancer, prostate cancer or other conditions as described in Example 13, below. The individuals is screened while they are asymptomatic to determine their risk of developing cancer, prostate cancer or other conditions at a subsequent time. Alternatively, individuals suffering from cancer, prostate cancer or other conditions is screened for the presence of polymorphisms in the PG1 gene or protein in order to determine whether therapies which target the PG1 gene or protein should be applied.

Example 14

Nucleic acid samples are obtained from a symptomatic or asymptomatic individual. The nucleic acid samples is obtained from blood cells as described above or is obtained from other tissues or organs. For individuals suffering from prostate cancer, the nucleic acid sample is obtained from the tumor. The nucleic acid sample may comprise DNA, RNA, or both. The nucleotides at positions in the PG1 gene where mutations lead to prostate cancer or other detectable phenotypes are determined for the nucleic acid sample.

In one embodiment, a PCR amplification is conducted on the nucleic acid sample as described above to amplify regions in which polymorphisms associated with prostate cancer or other detectable phenotypes have been identified. The amplification products are sequenced to determine whether the individual possesses one or more PG1 polymorphisms associated with prostate cancer or other detectable phenotypes.

Alternatively, the nucleic acid sample is subjected to microsequencing reactions as described above to determine whether the individual possesses one or more PG1 polymorphisms associated with prostate cancer or another detectable phenotype resulting from a mutation in the PG1 gene.

In another embodiment, the nucleic acid sample is contacted with one or more allele specific oligonucleotides which specifically hybridize to one or more PG1 alleles associated with prostate cancer or another detectable phenotype. The nucleic acid sample is also contacted with a second PG1 oligonucleotide capable of producing an amplification product when used with the allele specific
5 oligonucleotide in an amplification reaction. The presence of an amplification product in the amplification reaction indicates that the individual possesses one or more PG1 alleles associated with prostate cancer or another detectable phenotype.

Determination of PG1 Expression Levels

As discussed above, PG1 polymorphisms associated with cancer, prostate cancer or other
10 detectable phenotypes may exert their effects by increasing, decreasing, or eliminating PG1 expression, or in altering the frequency of various transcription species. Accordingly, PG1 expression levels in individuals suffering from cancer, prostate cancer or other detectable phenotypes is compared to those of unaffected individuals to determine whether over-expression, under-expression, loss of expression, or changes in the relative frequency of transcription species of PG1 causes cancer,
15 prostate cancer or another detectable phenotype. Individuals is tested to determine whether they are at risk of developing cancer, or prostate cancer at a subsequent time or whether they suffer from prostate cancer resulting from a mutation in the PG1 gene by determining whether they exhibit a level of PG1 expression associated with prostate cancer. Similarly, individuals is tested to determine whether they suffer from another PG1 mediated detectable phenotype or whether they are at risk of suffering from
20 such a condition at a subsequent time.

Expression levels in nucleic acid samples from affected and unaffected individuals is determined by performing Northern blots using detectable probes derived from the PG1 gene or the PG1 cDNA. A variety of conventional Northern blotting procedures is used to detect and quantitate PG1 expression and the frequencies of the various transcription species of PG1, including those
25 disclosed in Current Protocols in Molecular Biology, John Wiley 503 Sons, Inc. 1997 and Sambrook et al. Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press, 1989.

Alternatively, PG1 expression levels is determined as described in Example 15, below.

Example 15

30 Expression levels and patterns of PG1 is analyzed by solution hybridization with long probes as described in International Patent Application No. WO 97/05277. Briefly, the PG1 cDNA or the PG1 genomic DNA described above, or fragments thereof, is inserted at a cloning site immediately downstream of a bacteriophage (T3, T7 or SP6) RNA polymerase promoter to produce antisense RNA. Preferably, the PG1 insert comprises at least 100 or more consecutive nucleotides of the genomic DNA
35 sequence of SEQ ID NO: 1 or the cDNA sequences of SEQ ID NO: 3. The plasmid is linearized and transcribed in the presence of ribonucleotides comprising modified ribonucleotides (i.e. biotin-UTP and

DIG-UTP). An excess of this doubly labeled RNA is hybridized in solution with mRNA isolated from cells or tissues of interest. The hybridizations are performed under standard stringent conditions (40-50° C for 16 hours in an 80% formamide, 0.4 M NaCl buffer, pH 7-8). The unhybridized probe is removed by digestion with ribonucleases specific for single-stranded RNA (i.e. RNases CL3, T1, Phy M, U2 or A). The presence of the biotin-UTP modification enables capture of the hybrid on a microtitration plate coated with streptavidin. The presence of the DIG modification enables the hybrid to be detected and quantified by ELISA using an anti-DIG antibody coupled to alkaline phosphatase.

Quantitative analysis of PG1 gene expression may also be performed using arrays as described in Sections II and X. As used here, the term array means an arrangement of a plurality of nucleic acids of sufficient length to permit specific detection of expression of PG1 mRNAs capable of hybridizing thereto. For example, the arrays may contain a plurality of nucleic acids derived from genes whose expression levels are to be assessed. The arrays may include the PG1 genomic DNA of SEQ ID NO:179, the PG1 cDNA of SEQ ID NO:3 or the sequences complementary thereto or fragments thereof. The array may contain some or all of the known alternative splice or transcription species of PG1, including the species in SEQ ID NOs: 3, and 112-124 to determine the relative frequency of particular transcription species. Alternatively, the array may contain polynucleotides which overlap all of the potential splice junctions, including, for example SEQ ID NOs: 137-178, so that the frequency of particular splice junctions can be determined and correlated with traits or used in diagnostics just as expressions levels are. Preferably, the fragments are at least 15 nucleotides in length. In other embodiments, the fragments are at least 25 nucleotides in length. In some embodiments, the fragments are at least 50 nucleotides in length. More preferably, the fragments are at least 100 nucleotides in length. In another preferred embodiment, the fragments are more than 100 nucleotides in length. In some embodiments the fragments is more than 500 nucleotides in length.

For example, quantitative analysis of PG1 gene expression is performed with a complementary DNA microarray as described by Schena et al. (Science 270:467-470, 1995; Proc. Natl. Acad. Sci. U.S.A. 93:10614-10619, 1996). Full length PG1 cDNAs or fragments thereof are amplified by PCR and arrayed from a 96-well microtiter plate onto silylated microscope slides using high-speed robotics. Printed arrays are incubated in a humid chamber to allow rehydration of the array elements and rinsed, once in 0.2% SDS for 1 min, twice in water for 1 min and once for 5 min in sodium borohydride solution. The arrays are submerged in water for 2 min at 95° C, transferred into 0.2% SDS for 1 min, rinsed twice with water, air dried and stored in the dark at 25° C.

Cell or tissue mRNA is isolated or commercially obtained and probes are prepared by a single round of reverse transcription. Probes are hybridized to 1 cm² microarrays under a 14 x 14 mm glass coverslip for 6-12 hours at 60° C. Arrays are washed for 5 min at 25° C in low stringency wash buffer (1 x SSC/0.2% SDS), then for 10 min at room temperature in high stringency wash buffer (0.1 x SSC/0.2% SDS). Arrays are scanned in 0.1 x SSC using a fluorescence laser scanning device fitted with

a custom filter set. Accurate differential expression measurements are obtained by taking the average of the ratios of two independent hybridizations.

Quantitative analysis of PG1 gene expression may also be performed with full length PG1 cDNAs or fragments thereof in complementary DNA arrays as described by Pietu et al. (Genome Research 6:492-503, 1996). The full length PG1 cDNA or fragments thereof is PCR amplified and spotted on membranes. Then, mRNAs originating from various tissues or cells are labeled with radioactive nucleotides. After hybridization and washing in controlled conditions, the hybridized mRNAs are detected by phospho-imaging or autoradiography. Duplicate experiments are performed and a quantitative analysis of differentially expressed mRNAs is then performed.

Alternatively, expression analysis using the PG1 genomic DNA, the PG1 cDNA, or fragments thereof can be done through high density nucleotide arrays as described by Lockhart et al. (Nature Biotechnology 14: 1675-1680, 1996) and Sosnowsky et al. (Proc. Natl. Acad. Sci. 94:1119-1123, 1997). Oligonucleotides of 15-50 nucleotides from the sequences of the PG1 genomic DNA of SEQ ID NO: 179, the PG1 cDNA of SEQ ID NO: 3, 112-124 or the sequences complementary thereto, are synthesized directly on the chip (Lockhart et al., supra) or synthesized and then addressed to the chip (Sosnowski et al., supra).

PG1 cDNA probes labeled with an appropriate compound, such as biotin, digoxigenin or fluorescent dye, are synthesized from the appropriate mRNA population and then randomly fragmented to an average size of 50 to 100 nucleotides. The said probes are then hybridized to the chip. After washing as described in Lockhart et al., supra and application of different electric fields (Sosnowsky et al., Proc. Natl. Acad. Sci. 94:1119-1123), the dyes or labeling compounds are detected and quantified. Duplicate hybridizations are performed. Comparative analysis of the intensity of the signal originating from cDNA probes on the same target oligonucleotide in different cDNA samples indicates a differential expression of PG1 mRNA.

The above methods may also be used to determine whether an individual exhibits a PG1 expression pattern associated with cancer, prostate cancer or other diseases. In such methods, nucleic acid samples from the individual are assayed for PG1 expression as described above. If a PG1 expression pattern associated with cancer, prostate cancer, or another disease is observed, an appropriate diagnosis is rendered and appropriate therapeutic techniques which target the PG1 gene or protein is applied.

The above methods may also be applied using allele specific probes to determine whether an individual possesses a PG1 allele associated with cancer, prostate cancer, or another disease. In such approaches, one or more allele specific oligonucleotides containing polymorphic nucleotides in the PG1 gene which are associated with prostate cancer are fixed to a microarray. The array is contacted with a nucleic acid sample from the individual being tested under conditions which permit allele specific hybridization of the sample nucleic acid to the allele specific PG1 probes. Hybridization of the sample

nucleic acid to one or more of the allele specific PGI probes indicates that the individual suffers from prostate cancer caused by the PGI gene or that the individual is at risk for developing prostate cancer at a subsequent time. Alternatively, any of the genotyping methods described in Section X is utilized..

Use of the Biallelic Markers Of The Invention In Diagnostics

5 The biallelic markers of the present invention can also be used to develop diagnostics tests capable of identifying individuals who express a detectable trait as the result of a specific genotype or individuals whose genotype places them at risk of developing a detectable trait at a subsequent time.

10 The diagnostic techniques of the present invention may employ a variety of methodologies to determine whether a test subject has a biallelic marker pattern associated with an increased risk of developing a detectable trait or whether the individual suffers from a detectable trait as a result of a particular mutation, including methods which enable the analysis of individual chromosomes for haplotyping, such as family studies, single sperm DNA analysis or somatic hybrids. The trait analyzed using the present diagnostics is any detectable trait, cancer, prostate cancer or another disease, a response to an anti-cancer, or anti-prostate cancer, or side effects to an anti-cancer or anti-prostate cancer agent. Diagnostics, which analyze and predict response to a drug or side effects to a drug, is used to determine whether an individual should be treated with a particular drug. For example, if the diagnostic indicates a likelihood that an individual will respond positively to treatment with a particular drug, the drug is administered to the individual. Conversely, if the diagnostic indicates that an individual is likely to respond negatively to treatment with a particular drug, an alternative course of treatment is prescribed. A negative response is defined as either the absence of an efficacious response or the presence of toxic side effects.

15 Clinical drug trials represent another application for the markers of the present invention. One or more markers indicative of response to an anti-cancer or anti-prostate cancer agent or to side effects to an anti-cancer or anti-prostate cancer agent is identified using the methods described in Section XI, below. Thereafter, potential participants in clinical trials of such an agent is screened to identify those individuals most likely to respond favorably to the drug and exclude those likely to experience side effects. In that way, the effectiveness of drug treatment is measured in individuals who respond positively to the drug, without lowering the measurement as a result of the inclusion of individuals who are unlikely to respond positively in the study and without risking undesirable safety problems. Preferably, in such diagnostic methods, a nucleic acid sample is obtained from the individual and this sample is genotyped using methods described in Section X.

20 Another aspect of the present invention relates to a method of determining whether an individual is at risk of developing a trait or whether an individual expresses a trait as a consequence of possessing a particular trait-causing allele. The present invention relates to a method of determining whether an individual is at risk of developing a plurality of traits or whether an individual expresses a plurality of traits as a result of possessing a particular trait-causing allele. These methods involve

obtaining a nucleic acid sample from the individual and determining whether the nucleic acid sample contains one or more alleles of one or more biallelic markers indicative of a risk of developing the trait or indicative that the individual expresses the trait as a result of possessing a particular trait-causing allele.

- 5 As described herein, the diagnostics is based on a single biallelic marker or a group of biallelic markers.

VL ASSAYING THE PG1 PROTEIN FOR INVOLVEMENT IN RECEPTOR/LIGAND INTERACTIONS

- 10 The expressed PG1 protein or portion thereof is evaluated for involvement in receptor/ligand interactions as described in Example 16 below.

Example 16

- 15 The proteins encoded by the PG1 gene or a portion thereof may also be evaluated for their involvement in receptor/ligand interactions. Numerous assays for such involvement are familiar to those skilled in the art, including the assays disclosed in the following references: Chapter 7.28 (Measurement of Cellular Adhesion under Static Conditions 7.28.1-7.28.22) in Current Protocols in Immunology, J.E. Coligan et al. Eds. Greene Publishing Associates and Wiley-Interscience; Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160, 1989; Stoltzenberg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995; Gyuris et al., Cell 75:791-803, 1993.

- 20 For example, the proteins of the present invention may demonstrate activity as receptors, receptor ligands or inhibitors or agonists of receptor/ligand interactions. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their
25 ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses). Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

- 30 The PG1 protein or portions thereof described above is used in drug screening procedures to identify molecules which are agonists, antagonists, or inhibitors of PG1 activity. The PG1 protein or portion thereof used in such analyses is free in solution or linked to a solid support. Alternatively, PG1 protein or portions thereof can be expressed on a cell surface. The cell may naturally express the PG1 protein or portion thereof or, alternatively, the cell may express the PG1 protein or portion thereof from
35 an expression vector such as those described below.

In one method of drug screening, eukaryotic or prokaryotic host cells which are stably transformed with recombinant polynucleotides in order to express the PG1 protein or a portion thereof are used in conventional competitive binding assays or standard direct binding assays. For example, the formation of a complex between the PG1 protein or a portion thereof and the agent being tested is measured in
5 direct binding assays. Alternatively, the ability of a test agent to prevent formation of a complex between the PG1 protein or a portion thereof and a known ligand is measured.

Alternatively, the high throughput screening techniques disclosed in the published PCT application WO 84/03564, is used. In such techniques, large numbers of small peptides to be tested for PG1 binding activity are synthesized on a surface and affixed thereto. The test peptides are contacted
10 with the PG1 protein or a portion thereof, followed by a wash step. The amount of PG1 protein or portion thereof which binds to the test compound is quantitated using conventional techniques.

In some methods, PG1 protein or a portion thereof is fixed to a surface and contacted with a test compound. After a washing step, the amount of test compound which binds to the PG1 protein or portion thereof is measured.

15 In another approach, the three dimensional structure of the PG1 protein or a portion thereof may be determined and used for rational drug design.

Alternatively, the PG1 protein or a portion thereof is expressed in a host cell using expression vectors such as those described herein. The PG1 protein or portion thereof is an isotype which is associated with prostate cancer or an isotype which is not associated with prostate cancer. The cells
20 expressing the PG1 protein or portion thereof are contacted with a series of test agents and the effects of the test agents on PG1 activity are measured. Test agents which modify PG1 activity is employed in therapeutic treatments.

The above procedures may also be applied to evaluate mutant PG1 proteins responsible for a detectable phenotype.

25 Identification of Proteins which Interact with the PG1 Protein

Proteins which interact with the PG1 protein is identified as described in Example 17, below.

Example 17

Proteins which interact with the PG1 protein or a portion thereof, is identified using two hybrid systems such as the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech). As described
30 in the manual accompanying the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech), nucleic acids encoding the PG1 protein or a portion thereof, are inserted into an expression vector such that they are in frame with DNA encoding the DNA binding domain of the yeast transcriptional activator GAL4. cDNAs in a cDNA library which encode proteins which might interact with the polypeptides encoded by the nucleic acids encoding the PG1 protein or a portion thereof are inserted into a second
35 expression vector such that they are in frame with DNA encoding the activation domain of GAL4. The two expression plasmids are transformed into yeast and the yeast are plated on selection medium which

selects for expression of selectable markers on each of the expression vectors as well as GAL4 dependent expression of the HIS3 gene. Transformants capable of growing on medium lacking histidine are screened for GAL4 dependent lacZ expression. Those cells which are positive in both the histidine selection and the lacZ assay contain plasmids encoding proteins which interact with the polypeptide
5 encoded by the nucleic acid inserts.

Alternatively, the system described in Lustig et al., *Methods in Enzymology* 283: 83-99 (1997), is used for identifying molecules which interact with the PG1 protein or a portion thereof. In such systems, in vitro transcription reactions are performed on vectors containing an insert encoded the PG1 protein or a portion thereof cloned downstream of a promoter which drives in vitro transcription. The
10 resulting mRNA is introduced into *Xenopus laevis* oocytes. The oocytes are then assayed for a desired activity.

Alternatively, the in vitro transcription products produced as described above is translated in vitro. The in vitro translation products can be assayed for a desired activity or for interaction with a known polypeptide.

15 The system described in U.S. Patent No. 5,654,150 may also be used to identify molecules which interact with the PG1 protein or a portion thereof. In this system, pools of cDNAs are transcribed and translated in vitro and the reaction products are assayed for interaction with a known polypeptide or antibody.

Proteins or other molecules interacting with the PG1 protein or portions thereof can be found
20 by a variety of additional techniques. In one method, affinity columns containing the PG1 protein or a portion thereof can be constructed. In some versions of this method the affinity column contains chimeric proteins in which the PG1 protein or a portion thereof is fused to glutathione S-transferase. A mixture of cellular proteins or pool of expressed proteins as described above is applied to the affinity column. Proteins interacting with the polypeptide attached to the column can then be isolated
25 and analyzed on 2-D electrophoresis gel as described in Ramunsen et al. *Electrophoresis*, 18, 588-598 (1997). Alternatively, the proteins retained on the affinity column can be purified by electrophoresis based methods and sequenced. The same method can be used to isolate antibodies, to screen phage display products, or to screen phage display human antibodies.

Proteins interacting with the PG1 protein or portions thereof can also be screened by using an
30 Optical Biosensor as described in Edwards et Leatherbarrow, *Analytical Biochemistry*, 246, 1-6 (1997). The main advantage of the method is that it allows the determination of the association rate between the protein and other interacting molecules. Thus, it is possible to specifically select interacting molecules with a high or low association rate. Typically a target molecule is linked to the sensor surface (through a carboxymethyl dextran matrix) and a sample of test molecules is placed in
35 contact with the target molecules. The binding of a test molecule to the target molecule causes a change in the refractive index and/ or thickness. This change is detected by the Biosensor provided it

occurs in the evanescent field (which extend a few hundred nanometers from the sensor surface). In these screening assays, the target molecule can be the PG1 protein or a portion thereof and the test sample can be a collection of proteins extracted from tissues or cells, a pool of expressed proteins, combinatorial peptide and/ or chemical libraries, or phage displayed peptides. The tissues or cells
5 from which the test proteins are extracted can originate from any species.

In other methods, a target protein is immobilized and the test population is the PG1 protein or a portion thereof.

To study the interaction of the PG1 protein or a portion thereof with drugs, the microdialysis coupled to HPLC method described by Wang et al., *Chromatographia*, 44, 205-208(1997) or the
10 affinity capillary electrophoresis method described by Busch et al., *J. Chromatogr.* 777:311-328 (1997).

The above procedures may also be applied to evaluate mutant PG1 proteins responsible for a detectable phenotype.

VII. PRODUCTION OF ANTIBODIES AGAINST PG1 POLYPEPTIDES

15 Any PG1 polypeptide or whole protein (SEQ ID NOs: 4, 5, 70, 74, 125-136) whether human, mouse or mammalian is used to generate antibodies capable of specifically binding to expressed PG1 protein or fragments thereof as described in Example 16, below. The antibodies is capable of binding the full length PG1 protein. PG1 proteins which result from naturally occurring mutant, particularly functional mutants of PG1, including SEQ ID NO: 70, which may used in the
20 production of antibodies. The present invention also contemplates the use of polypeptides comprising a contiguous stretch of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 50, or 100 amino acids of any PG1 protein in the manufacture of antibodies. In a preferred embodiment the contiguous stretch of amino acids comprises the site of a mutation or functional mutation, including a deletion, addition, swap or truncation of the amino acids in the PG1
25 protein sequence. For instance, polypeptides that contain either the Arg and His residues at amino acid position 184, and polypeptides that contain either the Arg or Ile residue at amino acid position 293 of the SEQ ID NO: 4 in said contiguous stretch are particularly preferred embodiments of the invention and useful in the manufacture of antibodies to detect the presence and absence of these mutations. Similarly, polypeptides with a carboxy terminus at position 228 is a particularly preferred
30 embodiment of the invention and useful in the manufacture of antibodies to detect the presence and absence of the mutation shown in SEQ ID NOs: 69 and 70. Similarly, polypeptides that contain an peptide sequences of 8, 10, 12, 15, or 25 amino acids encoded over a naturally-occurring splice junction (the point at which two human PG1 exon (SEQ ID NOs: 100-111) are covalently linked) in said contiguous stretch are particularly preferred embodiments and useful in the manufacture of
35 antibodies to detect the presence, localization, and quantity of the various protein products of the PG1 alternative splice species.

Alternatively, the antibodies is screened so as to isolate those which are capable of binding an epitope-containing fragment of at least 8, 10, 12, 15, 20, 25, or 30 amino acids of a human, mouse, or mammalian PG1 protein, preferably a sequence selected from SEQ ID NOs: 4, 5, 70, 74, or 125-136.

Antibodies may also be generated which are capable of specifically binding to a given isoform
5 of the PG1 protein. For example, the antibodies is capable of specifically binding to an isoform of the PG1 protein which causes prostate cancer or another detectable phenotype which has been obtained as described above and expressed from an expression vector as described above. Alternatively, the antibodies is capable of binding to an isoform of the PG1 protein which does not cause prostate cancer. Such antibodies is used in diagnostic assays in which protein samples from an individual are evaluated
10 for the presence of an isoform of the PG1 protein which causes cancer or another detectable phenotype using techniques such as Western blotting or ELISA assays.

Non-human animals or mammals, whether wild-type or transgenic, which express a different species of PG1 than the one to which antibody binding is desired, and animals which do not express PG1 (i.e. an PG1 knock out animal as described in Section VIII.) are particularly useful for preparing
15 antibodies. PG1 knock out animals will recognize all or most of the exposed regions of PG1 as foreign antigens, and therefore produce antibodies with a wider array of PG1 epitopes. The humoral immune system of animals which produce a species of PG1 that resembles the antigenic sequence will preferentially recognize the differences between the animal's native PG1 species and the antigen sequence, and produce antibodies to these unique sites in the antigen sequence.

20

Example 18

Substantially pure protein or polypeptide is isolated from transfected or transformed cells containing an expression vector encoding the PG1 protein or a portion thereof as described in Example 11. The concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the
25 protein can then be prepared as follows:

A. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes in the PG1 protein or a portion thereof can be prepared from murine hybridomas according to the classical method of Kohler, G. and Milstein, C., Nature 256:495 (1975) or derivative methods thereof. Also see Harlow, E., and D. Lane. 1988. Antibodies A
30 Laboratory Manual. Cold Spring Harbor Laboratory. pp. 53-242.

Briefly, a mouse is repetitively inoculated with a few micrograms of the PG1 protein or a portion thereof over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising
35 aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are

identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, E., *Meth. Enzymol.* 70:419 (1980), and derivative methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. et al. *Basic Methods in Molecular Biology* Elsevier, New York. Section 21-2.

B. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogeneous epitopes in the PG1 protein or a portion thereof can be prepared by immunizing suitable non-human animal with the PG1 protein or a portion thereof, which can be unmodified or modified to enhance immunogenicity. A suitable non-human animal is preferably a non-human mammal is selected, usually a mouse, rat, rabbit, goat, or horse. Alternatively, a crude preparation which has been enriched for PG1 concentration can be used to generate antibodies. Such proteins, fragments or preparations are introduced into the non-human mammal in the presence of an appropriate adjuvant (e.g. aluminum hydroxide, RIBI, etc.) which is known in the art. In addition the protein, fragment or preparation can be pretreated with an agent which will increase antigenicity, such agents are known in the art and include, for example, methylated bovine serum albumin (mBSA), bovine serum albumin (BSA), Hepatitis B surface antigen, and keyhole limpet hemocyanin (KLH). Serum from the immunized animal is collected, treated and tested according to known procedures. If the serum contains polyclonal antibodies to undesired epitopes, the polyclonal antibodies can be purified by immunoaffinity chromatography.

Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. Techniques for producing and processing polyclonal antisera are known in the art, see for example, Mayer and Walker (1987). An effective immunization protocol for rabbits can be found in Vaitukaitis, J. et al. *J. Clin. Endocrinol. Metab.* 33:988-991 (1971).

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. et al., Chap. 19 in: *Handbook of Experimental Immunology* D. Wier (ed) Blackwell (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about $12 \cdot M$). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in: *Manual of Clinical Immunology*, 2d Ed. (Rose and Friedman, Eds.) Amer. Soc. For Microbiol., Washington, D.C. (1980).

Antibody preparations prepared according to either the monoclonal or the polyclonal protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances

in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

VIII. VECTORS AND THE USES OF POLYNUCLEOTIDES IN CELLS, ANIMALS, AND

5 HUMANS

The nucleic acids of the invention include expression vectors, amplification vectors, PCR-suitable polynucleotide primers, and vectors which are suitable for the introduction of a polynucleotide of the invention into an embryonic stem cells for the production of transgenic non-human animals. In addition, vectors which are suitable for the introduction of a polynucleotide of the invention into cells, organs and individuals, including human individuals, for the purposes of gene therapy to reduce the severity of or prevent genetic diseases associated with functional mutations in PG1 genes are encompassed by the present invention. Functional mutations in PG1 genes which are suitable as targets for the gene therapy and transgenic vectors and methods of the invention include, but are not limited to, mutations in the coding region of the PG1 gene which affect the amino acid sequence of the PG1 gene's product, mutations in the promoter or other regulatory regions which affect the levels of PG1 expression, mutations in the PG1 splice sites which affect length of the PG1 gene product or the relative frequency of PG1 alternative splicing species, and any other mutation which in any way affects the level or quality of PG1 expression or activity. The gene therapy methods can be achieved by targeting vectors and method for changing a mutant PG1 gene into a wild-type PG1 gene in a embryonic stem cell or somatic cell. Alternatively, the present invention also encompasses methods and vectors for introducing the expression of wild-type PG1 sequences without the disruption of any mutant PG1 which already reside in the cell, organ or individual.

The invention also embodies amplification vectors, which comprise a polynucleotide of the invention, and an origin of replication. Preferably, such amplification vectors further comprise restriction endonuclease sites flanking the polynucleotide, so as to facilitate cleavage and purification of the polynucleotides from the remainder of the amplification vector, and a selectable marker, so as to facilitate amplification of the amplification vector. Most preferably, the restriction endonuclease sites in the amplification vector are situated such that cleavage at those site would result in no other amplification vector fragments of a similar size.

Thus, such an amplification vector is transfected into a host cell compatible with the origin of replication of said amplification vector, wherein the host cell is a prokaryotic or eukaryotic cell, preferably a mammalian, insect, yeast, or bacterial cell, most preferably an Escherichia coli cell. The resulting transfected host cells is grown by culture methods known in the art, preferably under selection compatible with the selectable marker (e.g., antibiotics). The amplification vectors can be isolated and purified by methods known in the art (e.g., standard plasmid prep procedures). The polynucleotide of the invention can be cleaved with restriction enzymes that specifically cleave at the

restriction endonuclease sites flanking the polynucleotide, and the double-stranded polynucleotide fragment purified by techniques known in the art, including gel electrophoresis.

Alternatively linear polynucleotides comprising a polynucleotide of the invention is amplified by PCR. The PCR method is well known in the art and described in, e.g., U.S. Patent Nos. 4,683,195
5 and 4,683,202 and Saiki, R et al. 1988. Science 239:487-491, and European patent applications 86302298.4, 86302299.2 and 87300203.4, as well as Methods in Enzymology 1987 155:335-350.

The polynucleotides of the invention can also be derivatized in various ways, including those appropriate for facilitating transfection and/or gene therapy. The polynucleotides can be derivatized by attaching a nuclear localization signal to it to improve targeted delivery to the nucleus. One well-
10 characterized nuclear localization signal is the heptapeptide PKKKRKV (pro-lys-lys-lys-arg-lys-val). Preferably, in the case of polynucleotides in the form of a closed circle, the nuclear localization signal is attached via a modified loop nucleotide or spacer that forms a branching structure.

If it is to be used in vivo, the polynucleotide of the invention is derivatized to include ligands and/or delivery vehicles which provide dispersion through the blood, targeting to specific cell types,
15 or permit easier transit of cellular barriers. Thus, the polynucleotides of the invention is linked or combined with any targeting or delivery agent known in the art, including but not limited to, cell penetration enhancers, lipofectin, liposomes, dendrimers, DNA intercalators, and nanoparticles. In particular, nanoparticles for use in the delivery of the polynucleotides of the invention are particles of less than about 50 nanometers diameter, nontoxic, non-antigenic, and comprised of albumin and
20 surfactant, or iron as in the nanoparticle particle technology of SynGenix. In general the delivery vehicles used to target the polynucleotides of the invention may further comprise any cell specific or general targeting agents known in the art, and will have a specific trapping efficiency to the target cells or organs of from about 5 to about 35%.

The polynucleotides of the invention is used ex vivo in a gene therapy method for obtaining
25 cells or organs which produce wild-type PG1 or PG1 proteins which have been selectively mutated. The cells are created by incubation of the target cell with one or more of the above-described polynucleotides under standard conditions for uptake of nucleic acids, including electroporation or lipofection. In practicing an ex vivo method of treating cells or organs, the concentration of polynucleotides of the invention in a solution prepare to treat target cells or organs is from about 0.1
30 to about 100 μ M, preferably 0.5 to 50 μ M, most preferably from 1 to 10 μ M.

Alternatively, the oligonucleotides can be modified or co-administered for targeted delivery to the nucleus. Improved oligonucleotide stability is expected in the nucleus due to: (1) lower levels of DNases and RNases; and (2) higher oligonucleotide concentrations due to lower total volume.

Alternatively, the polynucleotides of the invention can be covalently bonded to biotin to form
35 a biotin-polynucleotide prodrug by methods known in the art, and co-administered with a receptor ligand bound to avidin or receptor specific antibody bound to avidin, wherein the receptor is capable

of causing uptake of the resulting polynucleotide-biotin-avidin complex into the cells. Receptors that cause uptake are known to those of skill in the art.

The invention encompasses vectors which are suitable for the introduction of a polynucleotide of the invention into an embryonic stem cell for the production of transgenic non-human animals, which in turn result in the expression of recombinant PG1 in the transgenic animal. Any appropriate vector system can be used for the introduction and expression of PG1 in transgenic animals, including for example yeast artificial chromosomes (YAC), bacterial artificial chromosomes (BAC), bacteriophage P1, and other vectors known in the art which are able to accommodate sufficiently large inserts to encode the PG1 protein or desired fragments thereof. Selected alterations, additions and deletions in the PG1 gene may optionally be achieved by site-directed mutagenesis. Once an appropriate vector system is chosen, the site-directed mutagenesis process may then be conducted by techniques well known in the art, and the fragment be returned and ligated to the larger vector from which it was cleaved. For site directed mutagenesis methods see, for example, Kunkel, T. 1985. Proc. Natl. Acad. Sci. U.S.A. 82:488; Bandeyar, M. et al. 1988. Gene 65: 129-133; Nelson, M., and M. McClelland 1992. Methods Enzymol. 216:279-303; Weiner, M. 1994. Gene 151: 119-123; Costa, G. and M. Weiner. 1994. Nucleic Acids Res. 22: 2423; Hu, G. 1993. DNA and Cell Biology 12:763-770; and Deng, W. and J. Nickoff. 1992. Anal. Biochem. 200:81.

Briefly, the transgenic technology used herein involves the inactivation, addition or replacement of a portion of the PG1 gene or the entire gene. For example the present technology includes the addition of PG1 genes with or without the inactivation of the non-human animal's native PG1 genes, as described in the preceding two paragraphs and in the Examples. The invention also encompasses the use of vectors, and the vectors themselves which target and modify an existing human PG1 gene in a stem cell, whether it is contained in a non-human animal cell where it was previously introduced into the germ line by transgenic technology or it is a native PG1 gene in a human pluripotent or somatic cell. This transgene technology usually relies on homologous recombination in a pluripotent cell that is capable of differentiating into germ cell tissue. A DNA construct that encodes an altered region of the non-human animal's PG1 gene that contains, for instance a stop codon to destroy expression, is introduced into the nuclei of embryonic stem cells. Preferably mice are used for this transgenic work. In a portion of the cells, the introduced DNA recombines with the endogenous copy of the cell's gene, replacing it with the altered copy. Cells containing the newly engineered genetic alteration are injected in a host embryo of the same species as the stem cell, and the embryo is reimplanted into a recipient female. Some of these embryos develop into chimeric individuals that possess germ cells entirely derived from the mutant cell line. Therefore, by breeding the chimeric progeny it is possible to obtain a new strain containing the introduced genetic alteration. See Capecchi 1989. Science. 244:1288-1292 for a review of this procedure.

The present invention encompasses the polynucleotides described herein, as well as the methods for making these polynucleotides including the method for creating a mutation in a human PG1 gene. In addition, the present invention encompasses cells which comprise the polynucleotides of the invention, including but not limited to amplification host cells comprising amplification vectors of the invention. Furthermore the present invention comprises the embryonic stem cells and transgenic non-human animals and mammals described herein which comprise a gene encoding a human PG1 protein.

DNA construct that enables directing temporal and spatial gene expression in recombinant host cells and in transgenic animals

10 In order to study the physiological and phenotype consequences of a lack of synthesis of the PG1 protein, both at the cellular level and at the multi-cellular organism level, in particular as regards to disorders related to abnormal cell proliferation, notably cancers, the invention also encompasses DNA constructs and recombinant vectors enabling a conditional expression of a specific allele of the PG1 genomic sequence or cDNA and also of a copy of this genomic sequence or cDNA harboring
15 substitutions, deletions, or additions of one or more bases as regards to the PG1 nucleotide sequence of SEQ ID NOs: 3, 112-125, 179, 182-184, or a fragment thereof, these base substitutions, deletions or additions being located either in an exon, an intron or a regulatory sequence, but preferably in a 5'-regulatory sequence of a mammalian PG1 gene, more preferably SEQ ID NO: 180 or in an exon of the PG1 genomic sequence or within the PG1 cDNA of SEQ ID NOs 3, 112-125, or 184.

20 A first preferred DNA construct is based on the tetracycline resistance operon tet from E. coli transposon Tn110 for controlling the PG1 gene expression, such as described by Gossen M. et al., 1992, Proc. Natl. Acad. Sci. USA, 89: 5547-5551; Gossen M. et al., 1995, Science, 268: 1766-1769; and Furth P.A. et al., 1994, Proc. Natl. Acad. Sci. USA, 91: 9302-9306. Such a DNA construct contains seven tet operator sequences from Tn10 (tetop) that are fused to either a minimal promoter or
25 a 5'-regulatory sequence of the PG1 gene, said minimal promoter or said PG1 regulatory sequence being operably linked to a polynucleotide of interest that codes either for a sense or an antisense oligonucleotide or for a polypeptide, including a PG1 polypeptide or a peptide fragment thereof. This DNA construct is functional as a conditional expression system for the nucleotide sequence of interest when the same cell also comprises a nucleotide sequence coding for either the wild type (tTA) or the
30 mutant (rTA) repressor fused to the activating domain of viral protein VP16 of herpes simplex virus, placed under the control of a promoter, such as the HCMVIE1 enhancer/promoter or the MMTV-LTR. Indeed, a preferred DNA construct of the invention will comprise both the polynucleotide containing the tet operator sequences and the polynucleotide containing a sequence coding for the tTA or the rTA repressor.

In the specific embodiment wherein the conditional expression DNA construct contains the sequence encoding the mutant tetracycline repressor rTA, the expression of the polynucleotide of interest is silent in the absence of tetracycline and induced in its presence.

DNA constructs allowing homologous recombination : replacement vectors

5 A second preferred DNA construct will comprise, from 5'-end to 3'-end : (a) a first nucleotide sequence that is comprised of a PG1 sequence preferably a PG1 genomic sequence; (b) a nucleotide sequence comprising a positive selection marker, such as the marker for neomycin resistance (neo); and (c) a second nucleotide sequence that comprised of a PG1 sequence preferably a PG1 genomic sequence, and is located on the genome downstream the first PG1 nucleotide sequence (a).

10 In a preferred embodiment, this DNA construct also comprises a negative selection marker located upstream the nucleotide sequence (a) or downstream the nucleotide sequence (b). Preferably, the negative selection marker consists of the thymidine kinase (tk) gene (Thomas K.R. et al., 1986, Cell, 44: 419-428), the hygromycin beta gene (Te Riele et al., 1990, Nature, 348: 649-651), the hprt gene (Van der Lugt et al., 1991, Gene, 105: 263-267; and Reid L.H. et al., 1990, Proc. Natl. Acad. Sci. USA, 87: 4299-4303) or the Diphtheria toxin A fragment (Dt-A) gene (Nada S. et al., 1993, Cell, 73: 1125-1135; Yagi T. et al., 1990, Proc. Natl. Acad. Sci. USA, 87: 9918-9922). Preferably, the positive selection marker is located within a PG1 exon sequence so as to interrupt the sequence encoding a PG1 protein.

20 These replacement vectors are described for example by Thomas K.R. et al., 1986, Cell, 44: 419-428; Thomas K.R. et al., 1987, Cell, 51: 503-512; Mansour S.L. et al., 1988, Nature, 336: 348-352; and Koller et al., 1992, Annu. Rev. Immunol., 10: 705-30.

The first and second nucleotide sequences (a) and (c) is located at any point within a PG1 regulatory sequence, an intronic sequence, an exon sequence or a sequence containing both regulatory and/or intronic and/or exon sequences. The length of nucleotide sequences (a) and (c) is determined empirically by one of ordinary skill in the art. Nucleotide sequences (a) and (c) or any length are specifically contemplated in the present invention, however, lengths ranging from 1 kb to 50 kb, preferably from 1 kb to 10 kb, more preferably from 2 kb to 6 kb and most preferably from 2 kb to 4 kb are normally used.

DNA constructs allowing homologous recombination : Cre-loxP system.

30 These new DNA constructs make use of the site-specific recombination system of the P1 phage. The P1 phage possesses a recombinase called Cre which interacts specifically with a 34 base pairs loxP site. The loxP site is composed of two palindromic sequences of 13 bp separated by a 8 bp conserved sequence (Hoess et al., 1986, Nucleic Acids Res., 14: 2287-2300). The recombination by the Cre enzyme between two loxP sites having an identical orientation leads to the deletion of the DNA fragment.

35

The Cre-loxP system used in combination with a homologous recombination technique has been first described by Gu H. et al., 1993, Cell, 73: 1155-1164 ; and Gu H. et al., 1994, Science, 265: 103-106. Briefly, a nucleotide sequence of interest to be inserted in a targeted location of the genome harbors at least two loxP sites in the same orientation and located at the respective ends of a nucleotide sequence to be excised from the recombinant genome. The excision event requires the presence of the recombinase (Cre) enzyme within the nucleus of the recombinant host cell. The recombinase enzyme is brought at the desired time either by (a) incubating the recombinant host cells in a culture medium containing this enzyme, by injecting the Cre enzyme directly into the desired cell, such as described by Araki K. et al., 1995, Proc. Natl. Acad. Sci. USA, 92: 160-164 ; or by lipofection of the enzyme into the cells, such as described by Baubonis et al., 1993, Nucleic Acids Res., 21: 2025-2029; (b) transfecting the cell host with a vector comprising the Cre coding sequence operably linked to a promoter functional in the recombinant cell host, which promoter being optionally inducible, said vector being introduced in the recombinant cell host, such as described by Gu H. et al., 1993, Cell, 73: 1155-1164; and Sauer B. et al., 1988, Proc. Natl. Acad. Sci. USA, 85: 5166-5170; (c) introducing in the genome of the host cell a polynucleotide comprising the Cre coding sequence operably linked to a promoter functional in the recombinant cell host, which promoter is optionally inducible, and said polynucleotide being inserted in the genome of the cell host either by a random insertion event or an homologous recombination event, such as described by Gu H. et al., 1994, Science, 265: 103-106.

In the specific embodiment wherein the vector containing the sequence to be inserted in the PG1 gene by homologous recombination is constructed in such a way that selectable markers are flanked by loxP sites of the same orientation, it is possible, by treatment by the Cre enzyme, to eliminate the selectable markers while leaving the PG1 sequences of interest that have been inserted by an homologous recombination event. Again, two selectable markers are needed: a positive selection marker to select for the recombination event and a negative selection marker to select for the homologous recombination event. Vectors and methods using the Cre-loxP system are described by Zou Y.R. et al., 1994, Curr. Biol., 4: 1099-1103.

Thus, a third preferred DNA construct of the invention comprises, from 5'-end to 3'-end: (a) a first nucleotide sequence that is comprised of a PG1 sequence, preferably a PG1 genomic sequence; (b) a nucleotide sequence comprising a polynucleotide encoding a positive selection marker, such as the marker for neomycin resistance (neo), said nucleotide sequence comprising additionally two sequences defining a site recognized by a recombinase, such as a loxP site, the two sites being placed in the same orientation; and (c) a second nucleotide sequence that is comprised of a PG1 sequence, preferably a PG1 genomic sequence, and is located on the genome downstream of the first PG1 nucleotide sequence (a).

The sequences defining a site recognized by a recombinase, such as a loxP site, are preferably located within the nucleotide sequence (b) at suitable locations bordering the nucleotide sequence for

which the conditional excision is sought. In one specific embodiment, two loxP sites are located at each side of the positive selection marker sequence, in order to allow its excision at a desired time after the occurrence of the homologous recombination event.

In a preferred embodiment of a method using the third DNA construct described above, the excision of the polynucleotide fragment bordered by the two sites recognized by a recombinase, preferably two loxP sites, is performed at a desired time, due to the presence within the genome of the recombinant host cell of a sequence encoding the Cre enzyme operably linked to a promoter sequence, preferably an inducible promoter, more preferably a tissue-specific promoter sequence and most preferably a promoter sequence which is both inducible and tissue-specific, such as described by Gu H. et al., 1994, Science, 265: 103-106.

The presence of the Cre enzyme within the genome of the recombinant cell host may result of the breeding of two transgenic animals, the first transgenic animal bearing the PG1-derived sequence of interest containing the loxP sites as described above and the second transgenic animal bearing the Cre coding sequence operably linked to a suitable promoter sequence, such as described by Gu H. et al., 1994, Science, 265: 103-106. Spatio-temporal control of the Cre enzyme expression may also be achieved with an adenovirus based vector that contains the Cre gene thus allowing infection of cells, or in vivo infection of organs, for delivery of the Cre enzyme, such as described by Anton M. et al., 1995, J. Virol., 69: 4600-4606; and Kanegae Y. et al., 1995, Nucl. Acids Res., 23: 3816-3821.

The DNA constructs described above is used to introduce a desired nucleotide sequence of the invention, preferably a PG1 genomic sequence or a PG1 cDNA sequence, and most preferably an altered copy of a PG1 genomic or cDNA sequence, within a predetermined location of the targeted genome, leading either to the generation of an altered copy of a targeted gene (knock-out homologous recombination) or to the replacement of a copy of the targeted gene by another copy sufficiently homologous to allow an homologous recombination event to occur (knock-in homologous recombination).

Nuclear antisense DNA constructs

Preferably, the antisense polynucleotides of the invention have a 3' polyadenylation signal that has been replaced with a self-cleaving ribozyme sequence, such that RNA polymerase II transcripts are produced without poly(A) at their 3' ends, these antisense polynucleotides being incapable of export from the nucleus, such as described by Liu Z. et al., 1994, Proc. Natl. Acad. Sci. USA, 91: 4528-4262. In a preferred embodiment, these PG1 antisense polynucleotides also comprise, within the ribozyme cassette, a histone stem-loop structure to stabilize cleaved transcripts against 3'-5' exonucleolytic degradation, such as described by Eckner R. et al., 1991, EMBO J., 10: 3513-3522.

Expression Vectors

The polynucleotides of the invention also include expression vectors. Expression vector systems, control sequences and compatible host are known in the art. For a review of these systems

see, for example, U.S. Patent No. 5,350,671, columns 45-48. Any of the standard methods known to those skilled in the art for the insertion of DNA fragments into a vector is used to construct expression vectors containing a chimeric gene consisting of appropriate transcriptional/translational control signals and the protein coding sequences. These methods may include in vitro recombinant DNA and
5 synthetic techniques and in vivo recombinants (genetic recombination).

Expression of a polypeptide, peptide or derivative, or analogs thereof encoded by a polynucleotide sequence in SEQ ID NOs: 3, 69, 100-112, or 179-184 is regulated by a second nucleic acid sequence so that the protein or peptide is expressed in a host transformed with the recombinant DNA molecule. For example, expression of a protein or peptide is controlled by any
10 promoter/enhancer element known in the art. Promoters which is used to control expression include, but are not limited to, the CMV promoter, the SV40 early promoter region (Bernoist and Chambon, 1981, *Nature* 290:304-310), the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (Yamamoto, et al., 1980, *Cell* 22:787-797), the herpes thymidine kinase promoter (Wagner et al., 1981, *Proc. Natl. Acad. Sci. U.S.A.* 78:1441-1445), the regulatory sequences of the
15 metallothionein gene (Brinster et al., 1982, *Nature* 296:39-42); prokaryotic expression vectors such as the beta-lactamase promoter (Villa-Kamaroff, et al., 1978, *Proc. Natl. Acad. Sci. U.S.A.* 75:3727-3731), or the tac promoter (DeBoer, et al., 1983, *Proc. Natl. Acad. Sci. U.S.A.* 80:21-25); see also "Useful proteins from recombinant bacteria" in *Scientific American*, 1980, 242:74-94; plant expression vectors comprising the nopaline synthetase promoter region (Herrera-Estrella et al., 1983,
20 *Nature* 303:209-213) or the cauliflower mosaic virus 35S RNA promoter (Gardner, et al., 1981, *Nucl. Acids Res.* 9:2871), and the promoter of the photosynthetic enzyme ribulose biphosphate carboxylase (Herrera-Estrella et al., 1984, *Nature* 310:115-120); promoter elements from yeast or other fungi such as the Gal 4 promoter, the ADC (alcohol dehydrogenase) promoter, PGK (phosphoglycerol kinase) promoter, alkaline phosphatase promoter, and the following animal transcriptional control regions,
25 which exhibit tissue specificity and have been utilized in transgenic animals: elastase I gene control region which is active in pancreatic acinar cells (Swift et al., 1984, *Cell* 38:639-646; Ornitz et al., 1986, *Cold Spring Harbor Symp. Quant. Biol.* 50:399-409; MacDonald, 1987, *Hepatology* 7:425-515); insulin gene control region which is active in pancreatic beta cells (Hanahan, 1985, *Nature* 315:115-122), immunoglobulin gene control region which is active in lymphoid cells (Grosschedl et
30 al., 1984, *Cell* 38:647-658; Adames et al., 1985, *Nature* 318:533-538; Alexander et al., 1987, *Mol. Cell. Biol.* 7:1436-1444), mouse mammary tumor virus control region which is active in testicular, breast, lymphoid and mast cells (Leder et al., 1986, *Cell* 45:485-495), albumin gene control region which is active in liver (Pinkert et al., 1987, *Genes and Devel.* 1:268-276), alpha-fetoprotein gene control region which is active in liver (Krumlauf et al., 1985, *Mol. Cell. Biol.* 5:1639-1648; Hammer
35 et al., 1987, *Science* 235:53-58; alpha 1-antitrypsin gene control region which is active in the liver (Kelsey et al., 1987, *Genes and Devel.* 1:161-171), beta-globin gene control region which is active in

myeloid cells (Mogram et al., 1985, *Nature* 315:338-340; Kollias et al., 1986, *Cell* 46:89-94; myelin basic protein gene control region which is active in oligodendrocyte cells in the brain (Readhead et al., 1987, *Cell* 48:703-712); myosin light chain-2 gene control region which is active in skeletal muscle (Sani, 1985, *Nature* 314:283-286), and gonadotropic releasing hormone gene control region
5 which is active in the hypothalamus (Mason et al., 1986, *Science* 234:1372-1378).

Other suitable vectors, particularly for the expression of genes in mammalian cells, is selected from the group of vectors consisting of P1 bacteriophages, and bacterial artificial chromosomes (BACs). These types of vectors may contain large inserts ranging from about 80-90 kb (P1 bacteriophage) to about 300 kb (BACs).

10

P1 bacteriophage

The construction of P1 bacteriophage vectors such as p158 or p158/neo8 are notably described by Sternberg N.L., 1992, *Trends Genet.*, 8: 1-16; and Sternberg N.L., 1994, *Mamm. Genome*, 5: 397-404. Recombinant P1 clones comprising PG1 nucleotide sequences is designed for inserting large polynucleotides of more than 40 kb (Linton M.F. et al., 1993, *J. Clin. Invest.*, 92: 3029-
15 3037). To generate P1 DNA for transgenic experiments, a preferred protocol is the protocol described by McCormick et al., 1994, *Genet. Anal. Tech. Appl.*, 11: 158-164. Briefly, *E. coli* (preferably strain NS3529) harboring the P1 plasmid are grown overnight in a suitable broth medium containing 25 µg/ml of kanamycin. The P1 DNA is prepared from the *E. coli* by alkaline lysis using the Qiagen Plasmid Maxi kit (Qiagen, Chatsworth, CA, USA), according to the manufacturer's instructions. The
20 P1 DNA is purified from the bacterial lysate on two Qiagen-tip 500 columns, using the washing and elution buffers contained in the kit. A phenol/chloroform extraction is then performed before precipitating the DNA with 70% ethanol. After solubilizing the DNA in TE (10 mM Tris-HCl, pH 7.4, 1 mM EDTA), the concentration of the DNA is assessed by spectrophotometry.

When the goal is to express a P1 clone comprising PG1 nucleotide sequences in a transgenic
25 animal, typically in transgenic mice, it is desirable to remove vector sequences from the P1 DNA fragment, for example by cleaving the P1 DNA at rare-cutting sites within the P1 polylinker (SfiI, NotI or SalI). The P1 insert is then purified from vector sequences on a pulsed-field agarose gel, using methods similar using methods similar to those originally reported for the isolation of DNA from YACs (Schedl A. et al., 1993, *Nature*, 362: 258-261; and Peterson et al., 1993, *Proc. Natl. Acad. Sci.*
30 USA, 90: 7593-7597). At this stage, the resulting purified insert DNA can be concentrated, if necessary, on a Millipore Ultrafree-MC Filter Unit (Millipore, Bedford, MA, USA - 30,000 molecular weight limit) and then dialyzed against microinjection buffer (10 mM Tris-HCl, pH 7.4; 250 µM EDTA) containing 100 mM NaCl, 30 µM spermine, 70 µM spermidine on a microdialysis membrane (type VS, 0.025 µM from Millipore). The intactness of the purified P1 DNA insert is
35 assessed by electrophoresis on 1% agarose (Sea Kem GTG; FMC Bio-products) pulse-field gel and staining with ethidium bromide.

thereto. In some embodiments, the fragments may comprise more than 200 nucleotides of SEQ ID NOs: 3, 69, 112-125 or 184, or the sequence complementary thereto.

d) a non-native, purified or isolated nucleic acid comprising a nucleotide sequence selected from the group of SEQ ID NOs: 100 to 111, a sequence complementary thereto or a fragment or a
5 variant thereof.

e) a non-native, purified or isolated nucleic acid comprising a combination of at least two polynucleotides selected from the group consisting of SEQ ID NOs: 100 to 111, or the sequences complementary thereto wherein the polynucleotides are arranged within the nucleic acid, from the 5' end to the 3' end of said nucleic acid, in the same order than in SEQ NOs: 179, 182, or 183.

10 f) a non-native, purified or isolated nucleic acid comprising the nucleotide sequence SEQ ID NO: 180, or the sequences complementary thereto or a biologically active fragment or variant of the nucleotide sequence of SEQ ID NO: 180, or the sequence complementary thereto.

g) a non-native, purified or isolated nucleic acid comprising the nucleotide sequence SEQ ID NO: 181, or the sequence complementary thereto or a biologically active fragment or variant of the
15 nucleotide sequence of SEQ ID NO: 181 or the sequence complementary thereto.

h) a polynucleotide consisting of :

(1) a nucleic acid comprising a regulatory polynucleotide of SEQ ID NO: 180 or the sequences complementary thereto or a biologically active fragment or variant thereof

(2) a polynucleotide encoding a desired polypeptide or nucleic acid.

20 (3) Optionally, a nucleic acid comprising a regulatory polynucleotide of SEQ NO: 181, or the sequence complementary thereto or a biologically active fragment or variant thereof.

i) a DNA construct as described previously in the present specification.

The transgenic animals of the invention thus contain specific sequences of exogenous genetic material or "non-native" such as the nucleotide sequences described above in detail.

25 In a first preferred embodiment, these transgenic animals is good experimental models in order to study the diverse pathologies related to cell differentiation, in particular concerning the transgenic animals within the genome of which has been inserted one or several copies of a polynucleotide encoding a native PG1 protein, or alternatively a mutant PG1 protein.

In a second preferred embodiment, these transgenic animals may express a desired
30 polypeptide of interest under the control of the regulatory polynucleotides of the PG1 gene, leading to good yields in the synthesis of this protein of interest, and eventually a tissue specific expression of this protein of interest.

The design of the transgenic animals of the invention is made according to the conventional techniques well known from the one skilled in the art. For more details regarding the production of
35 transgenic animals, and specifically transgenic mice, it is referred to Sandou et al. (1994) and also to US Patents Nos 4,873,191, issued Oct.10, 1989, 5,464,764 issued Nov 7, 1995 and 5,789,215, issued

Aug 4, 1998.

Transgenic animals of the present invention are produced by the application of procedures which result in an animal with a genome that has incorporated exogenous genetic material. The procedure involves obtaining the genetic material, or a portion thereof, which encodes either a PG1 coding sequence, a PG1 regulatory polynucleotide or a DNA sequence encoding a PG1 antisense polynucleotide such as described in the present specification.

A recombinant polynucleotide of the invention is inserted into an embryonic or ES stem cell line. The insertion is preferably made using electroporation, such as described by Thomas K.R. et al., 1987, Cell, 51: 503-512. The cells subjected to electroporation are screened (e.g. by selection via selectable markers, by PCR or by Southern blot analysis) to find positive cells which have integrated the exogenous recombinant polynucleotide into their genome, preferably via an homologous recombination event. An illustrative positive-negative selection procedure that is used according to the invention is described by Mansour S.L. et al., 1988, Nature, 336: 348-352.

Then, the positive cells are isolated, cloned and injected into 3.5 days old blastocysts from mice, such as described by Bradley A., 1987, Production and analysis of chimaeric mice. In : E.J. Robertson (Ed.), Teratocarcinomas and embryonic stem cells : A practical approach. IRL Press, Oxford, pp.113. The blastocysts are then inserted into a female host animal and allowed to grow to term.

Alternatively, the positive ES cells are brought into contact with embryos at the 2.5 days old 8-16 cell stage (morulae) such as described by Wood S.A. et al., 1993, Proc. Natl. Acad. Sci. USA, 90: 4582-4585; or by Nagy A. et al., 1993, Proc. Natl. Acad. Sci. USA, 90: 8424-8428. The ES cells being internalized to colonize extensively the blastocyst including the cells which will give rise to the germ line. The offspring of the female host are tested to determine which animals are transgenic e.g. include the inserted exogenous DNA sequence and which are wild-type.

Thus, the present invention also concerns a transgenic animal containing a nucleic acid, a recombinant expression vector or a recombinant host cell according to the invention.

Recombinant cell lines derived from the transgenic animals of the invention.

A further object of the invention consists of recombinant host cells obtained from a transgenic animal described herein.

Recombinant cell lines is established in vitro from cells obtained from any tissue of a transgenic animal according to the invention, for example by transfection of primary cell cultures with vectors expressing onc-genes such as SV40 large T antigen, as described by Chou J.Y., 1989, Mol. Endocrinol., 3: 1511-1514 ; and Shay J.W. et al., 1991, Biochem. Biophys. Acta, 1072: 1-7.

commercially from a company specializing in custom oligonucleotide synthesis, such as GENSET, Paris, France.

The oligonucleotides is introduced into the cells using a variety of methods known to those skilled in the art, including but not limited to calcium phosphate precipitation, DEAE-Dextran, 5 electroporation, liposome-mediated transfection or native uptake.

Treated cells are monitored for altered cell function or reduced PG1 expression using techniques such as Northern blotting, RNase protection assays, or PCR based strategies to monitor the transcription levels of the PG1 gene in cells which have been treated with the oligonucleotide.

The oligonucleotides which are effective in inhibiting gene expression in tissue culture cells 10 may then be introduced in vivo using the techniques described above and in Example 19 at a dosage calculated based on the in vitro results, as described in Example 19.

In some embodiments, the natural (beta) anomers of the oligonucleotide units can be replaced with alpha anomers to render the oligonucleotide more resistant to nucleases. Further, an intercalating agent such as ethidium bromide, or the like, can be attached to the 3' end of the alpha oligonucleotide to 15 stabilize the triple helix. For information on the generation of oligonucleotides suitable for triple helix formation see Griffin et al. (Science 245:967-971 (1989)).

Alternatively, the PG1 cDNA, the PG1 genomic DNA, and the PG1 alleles of the present invention is used in gene therapy approaches in which expression of the PG1 protein is beneficial, as described in Example 21 below.

20 Example 21

The PG1 cDNA, the PG1 genomic DNA, and the PG1 alleles of the present invention may also be used to express the PG1 protein or a portion thereof in a host organism to produce a beneficial effect. In such procedures, the PG1 protein is transiently expressed in the host organism or stably expressed in the host organism. The expressed PG1 protein is used to treat conditions resulting from a lack of PG1 25 expression or conditions in which augmentation of existing levels of PG1 expression is beneficial.

A nucleic acid encoding the PG1 proteins of SEQ ID NO: 4, SEQ ID NO:5, or a PG1 allele is introduced into the host organism. The nucleic acid is introduced into the host organism using a variety of techniques known to those of skill in the art. For example, the nucleic acid is injected into the host organism as naked DNA such that the encoded PG1 protein is expressed in the host organism, thereby 30 producing a beneficial effect.

Alternatively, the nucleic acid encoding the PG1 proteins of SEQ ID NO: 4, SEQ ID NO: 5, or a PG1 allele is cloned into an expression vector downstream of a promoter which is active in the host organism. The expression vector is any of the expression vectors designed for use in gene therapy, including viral or retroviral vectors.

35 The expression vector is directly introduced into the host organism such that the PG1 protein is expressed in the host organism to produce a beneficial effect. In another approach, the expression vector

is introduced into cells in vitro. Cells containing the expression vector are thereafter selected and introduced into the host organism, where they express the PG1 protein to produce a beneficial effect.

IX. ISOLATION OF PG1 cDNA FROM NONHUMAN MAMMALS

The present invention encompasses mammalian PG1 sequences including genomic and cDNA sequences, as well as polypeptide sequences. The present invention also encompasses the use of PG1 genomic and cDNA sequences of the invention, including SEQ ID NOs: 179, 3, 182, and 183, in methods of isolating and characterizing PG1 nucleotide sequences derived from nonhuman mammals, in addition to sequences derived from human sequences. The human and mouse PG1 nucleic acid sequences of the invention can be used to construct primers and probes for amplifying and identifying PG1 genes in other nonhuman animals particularly mammals. The primers and probes used to identify nonhuman PG1 sequences is selected and used for the isolation of nonhuman PG1 utilizing the same techniques described above in Examples 4, 5, 6, 12 and 13.

In addition, sequence analysis of other homologous proteins is used to optimize the sequences of these primers and probes. As described above in the Analysis of the PG1 Protein Sequence, three boxes of homology were identified in the structure of the PG1 protein product when compared to proteins from a diverse range of organisms. See Figure 9. Using the assumption that the nucleotide sequences for these homologous proteins also show a high degree of homology, it is possible to construct primers that are specific for the PCR amplification of PG1 cDNA in nonhuman mammals.

Example 22

The primers BOXIed: AATCATCAAAGCACAGTTGACTGGAT (SEQ ID NO: 77) and BOXIIIer: ATAAACCACCGTAACATCATAAATTGCATCTAA (SEQ ID NO: 78) were designed as PCR primers from the human PG1 sequences after comparison with the sequence homologies of Figure 9. The BOXIed (SEQ ID NO: 77) and BOXIIIer (SEQ ID NO: 78) primers were used to amplify a mouse PG1 cDNA sequence from mouse liver marathon-ready cDNA (Clontech) under the conditions described above in Example 4. This PCR reaction yielded a product of approximately 400 base pairs, the boxI-boxIII fragment, which was subjected to automated dideoxy terminator sequencing and electrophoresed on ABI 377 sequencers as described above. Sequence analysis confirmed very high homology to human PG1 both at the nucleic acid and protein levels.

Primers were designed for RACE analysis using the 400 base pair boxI-boxIII fragment. Further sequence information was obtained using 5' and 3' RACE reactions on mouse liver marathon cDNA using two sets of these nested PCR primers: moPG1RACE5.350: AATCAAAAGCAACGTGAGTGGC (SEQ ID NO: 94) and moPG1RACE5.276: GCAAATGCCTGACTGGCTGA (SEQ ID NO: 93) for the 5' RACE reaction and moPG1RACE3.18: CTGCCAGACAGGATGCCCTA (SEQ ID NO: 90) and moPG1RACE3.63: ACAAGTTAAAATGGCTTCCGCTG (SEQ ID NO: 91) for the 3' RACE reaction. The PCR products of the RACE reactions were sequenced by primer walking using the following primers:

moPGrace3S473:	GAGATAAAAG ATAGGTTGCT CA	(SEQ ID NO: 79);
moPGrace3S526:	AAGAAACAAA TTTCCTGGG	(SEQ ID NO: 80);
moPGrace3S597:	TCTTGGGGAG TTTGACTG	(SEQ ID NO: 81);
moPGrace5R323:	GACCCCGGTG TAGTTCTC	(SEQ ID NO: 82);
5 moPGrace5R372:	CAGTAAAGCC GGTCTGTC	(SEQ ID NO: 83);
moPGrace5R444:	CAGGCCAGCA GGTAGGT	(SEQ ID NO: 84);
moPGrace5R492:	AGCAGGTAGC GCATAGAGT	(SEQ ID NO: 85).

Again a high degree of homology between the mouse sequence obtained from the primer walking and the human PG1 sequence was observed. An additional pair of nested primers were
 10 designed and utilized to further extend the 3' mouse PG1 sequence in yet another RACE reaction, moPG3RACE2: TGGGCACCTG GTTGTATGGA (SEQ ID NO: 95) and moPG3RACE2n: TCCTTGGCTG CCTGTGGTTT (SEQ ID NO: 96). The PCR product of this final RACE reaction was also sequenced by primer walking using the following primers:

moPG1RACE3R94:	CAAATGCATG TTGGCTGT	(SEQ ID NO: 92);
15 moPG3RACES20:	GATGGCTACA CATTGTATCA C	(SEQ ID NO: 97);
moPG3RACES5:	TCCTGAATTA AATAAGGAGT TTTC	(SEQ ID NO: 98);
moPG3RACES90:	GTTTGTATT AAAGCATAAG CAAG	(SEQ ID NO: 99).

The overlap in the 5' RACE, boxI-boxIII, and 3' RACE fragments allowed a single contiguous coding sequence for the mouse PG1 ortholog to be generated alignment of the three fragments. Primers
 20 were chosen from near the 5' and 3' ends of this predicted contiguous sequence (contig) in order to confirm the existence of such a transcript. PCR amplification was performed again on mouse liver marathon-ready cDNA (Clontech) with the chosen primers, moPG15: TGGCGAGCCGAGAGGATG (SEQ ID NO: 87) and moPG13LR2: GGAAACAATGTGATACAATGTGTAGCC (SEQ ID NO: 86) under the PCR conditions described above in Example 4. The resulting PCR product was a roughly 1.2
 25 kb DNA molecule and was shown to have an identical sequence to that of the deduced contig. Finally modified versions of the moPG15 and moPG13LR2 primers with the addition of EcoRI and BamHI sites, moPG15EcoRI: CGTGAATTCTGGCGAGCCGAGAGGATG (SEQ ID NO: 89) and moPG15BamI: CGTGGATCCGGAAACAATGTGATACAATGTGTAGCC (SEQ ID NO: 88) were used to obtain a PCR product that could be cloned into a pSKBluescript plasmid (Stratagene) cleaved
 30 with EcoRI and BamHI restriction enzymes. The mouse PG1 cDNA in the resulting construct was subjected to automated dideoxy terminator sequencing and electrophoresed on ABI 377 sequencers as described above. The sequence for mouse PG1 cDNA is reported in SEQ ID NO: 72, and the deduced amino acid sequence corresponding to the cDNA is reported in SEQ ID NO: 74.

Example 23

35 A mouse BAC library was constructed by the cloning of BamHI partially digested DNA of pluripotent embryonic stem cells, cell line ES-E14TG2a (ATCC CRL-1821) into pBeloBACII vector

plasmid. Approximately fifty-six thousand clones with an average inset size of 120 kb were picked individually and pooled for PCR screening as described above for human BAC library screening. These pools were screened with STS g34292 derived from the region of the mouse PG1 transcript corresponding to exon6 of the human gene. The upstream and downstream primers defining this STS
5 are: upstream amplification primer for g34292: ATTA AACAC GTACTGACAC CA (SEQ ID NO: 75), and downstream amplification primer for g34292: AGTCATGGAT GGTGGATTT (SEQ ID NO: 76). BAC C0281H06 tested positive for hybridizing to g34292. This BAC was isolated and sequenced by sub-cloning into pGenDel sequencing vector. The resulting partial genomic sequence for mouse PG1 is reported in SEQ ID NO: 73. This process was repeated and the resulting partial genomic sequences
10 for mouse PG1 is reported in SEQ ID NOs: 182 and 183.

Other mammalian PG1 cDNA and genomic sequences can be isolated by the methods of the present invention. PG1 genes in mammalian species have a region of at least 100, preferably 200, more preferably 500 nucleotides in each mammal's most abundant transcription species which has at least 75%, preferably 85%, more preferably 95% sequence homology to the most abundant human or
15 mouse cDNA species (SEQ ID NO: 3). PG1 proteins in mammalian species have a region of at least 40, preferably 90, more preferably 160 amino acids in the deduced amino acid sequence of the most abundant PG1 transcription species which has at least 75%, preferably 85%, more preferably 95% sequence homology to the deduced amino acid sequence of the most abundant human or mouse translations species (SEQ ID NO: 4 or 74).

20 **X. METHODS FOR GENOTYPING AN INDIVIDUAL FOR BIALLELIC MARKERS**

Methods are provided to genotype a biological sample for one or more biallelic markers of the present invention, all of which is performed in vitro. Such methods of genotyping comprise determining the identity of a nucleotide at an PG1-related biallelic marker by any method known in the art. These methods find use in genotyping case-control populations in association studies as well
25 as individuals in the context of detection of alleles of biallelic markers which, are known to be associated with a given trait, in which case both copies of the biallelic marker present in individual's genome are determined so that an individual is classified as homozygous or heterozygous for a particular allele.

These genotyping methods can be performed nucleic acid samples derived from a single
30 individual or pooled DNA samples.

Genotyping can be performed using similar methods as those described above for the identification of the biallelic markers, or using other genotyping methods such as those further described below. In preferred embodiments, the comparison of sequences of amplified genomic fragments from different individuals is used to identify new biallelic markers whereas
35 microsequencing is used for genotyping known biallelic markers in diagnostic and association study applications.

X.A. Source of DNA for genotyping

Any source of nucleic acids, in purified or non-purified form, can be utilized as the starting nucleic acid, provided it contains or is suspected of containing the specific nucleic acid sequence desired. DNA or RNA is extracted from cells, tissues, body fluids. As for the source of genomic DNA to be subjected to analysis, any test sample can be foreseen without any particular limitation. These test samples include biological samples, which can be tested by the methods of the present invention described herein, and include human and animal body fluids such as whole blood, serum, plasma, cerebrospinal fluid, urine, lymph fluids, and various external secretions of the respiratory, intestinal and genitourinary tracts, tears, saliva, milk, white blood cells, myelomas and the like; biological fluids such as cell culture supernatants; fixed tissue specimens including tumor and non-tumor tissue and lymph node tissues; bone marrow aspirates and fixed cell specimens. The preferred source of genomic DNA used in the present invention is from peripheral venous blood of each donor. Techniques to prepare genomic DNA from biological samples are well known to the skilled technician. While nucleic acids for use in the genotyping methods of the invention can be derived from any mammalian source, the test subjects and individuals from which nucleic acid samples are taken are generally understood to be human.

X.B. Amplification Of DNA Fragments Comprising Biallelic Markers

Methods and polynucleotides are provided to amplify a segment of nucleotides comprising one or more biallelic marker of the present invention. It will be appreciated that amplification of DNA fragments comprising biallelic markers is used in various methods and for various purposes and is not restricted to genotyping. Nevertheless, many genotyping methods, although not all, require the previous amplification of the DNA region carrying the biallelic marker of interest. Such methods specifically increase the concentration or total number of sequences that span the biallelic marker or include that site and sequences located either distal or proximal to it. Diagnostic assays may also rely on amplification of DNA segments carrying a biallelic marker of the present invention.

Amplification of DNA is achieved by any method known in the art. The established PCR (polymerase chain reaction) method or by developments thereof or alternatives. Amplification methods which can be utilized herein include but are not limited to Ligase Chain Reaction (LCR) as described in EP A 320 308 and EP A 439 182, Gap LCR (Wolcott, M.J., Clin. Microbiol. Rev. 5:370-386), the so-called "NASBA" or "3SR" technique described in Guatelli J.C. et al. (*Proc. Natl. Acad. Sci. USA* 87:1874-1878, 1990) and in Compton J. (*Nature* 350:91-92, 1991), Q-beta amplification as described in European Patent Application no 4544610, strand displacement amplification as described in Walker et al. (*Clin. Chem.* 42:9-13, 1996) and EP A 684 315 and, target mediated amplification as described in PCT Publication WO 9322461.

LCR and Gap LCR are exponential amplification techniques, both depend on DNA ligase to join adjacent primers annealed to a DNA molecule. In Ligase Chain Reaction (LCR), probe pairs are

used which include two primary (first and second) and two secondary (third and fourth) probes, all of which are employed in molar excess to target. The first probe hybridizes to a first segment of the target strand and the second probe hybridizes to a second segment of the target strand, the first and second segments being contiguous so that the primary probes abut one another in 5' phosphate-
5 3'hydroxyl relationship, and so that a ligase can covalently fuse or ligate the two probes into a fused product. In addition, a third (secondary) probe can hybridize to a portion of the first probe and a fourth (secondary) probe can hybridize to a portion of the second probe in a similar abutting fashion. Of course, if the target is initially double stranded, the secondary probes also will hybridize to the target complement in the first instance. Once the ligated strand of primary probes is separated from
10 the target strand, it will hybridize with the third and fourth probes which can be ligated to form a complementary, secondary ligated product. It is important to realize that the ligated products are functionally equivalent to either the target or its complement. By repeated cycles of hybridization and ligation, amplification of the target sequence is achieved. A method for multiplex LCR has also been described (WO 9320227). Gap LCR (GLCR) is a version of LCR where the probes are not adjacent
15 but are separated by 2 to 3 bases.

For amplification of mRNAs, it is within the scope of the present invention to reverse transcribe mRNA into cDNA followed by polymerase chain reaction (RT-PCR); or, to use a single enzyme for both steps as described in U.S. Patent No. 5,322,770 or, to use Asymmetric Gap LCR (RT-AGLCR) as described by Marshall R.L. et al. (*PCR Methods and Applications* 4:80-84, 1994).
20 AGLCR is a modification of GLCR that allows the amplification of RNA.

Some of these amplification methods are particularly suited for the detection of single nucleotide polymorphisms and allow the simultaneous amplification of a target sequence and the identification of the polymorphic nucleotide as it is further described in X.C.

The PCR technology is the preferred amplification technique used in the present invention. A
25 variety of PCR techniques are familiar to those skilled in the art. For a review of PCR technology, see Molecular Cloning to Genetic Engineering White, B.A. Ed. in *Methods in Molecular Biology* 67: Humana Press, Totowa (1997) and the publication entitled "PCR Methods and Applications" (1991, Cold Spring Harbor Laboratory Press). In each of these PCR procedures, PCR primers on either side of the nucleic acid sequences to be amplified are added to a suitably prepared nucleic acid sample along
30 with dNTPs and a thermostable polymerase such as Taq polymerase, Pfu polymerase, or Vent polymerase. The nucleic acid in the sample is denatured and the PCR primers are specifically hybridized to complementary nucleic acid sequences in the sample. The hybridized primers are extended. Thereafter, another cycle of denaturation, hybridization, and extension is initiated. The cycles are repeated multiple times to produce an amplified fragment containing the nucleic acid sequence between
35 the primer sites. PCR has further been described in several patents including US Patents 4,683,195, 4,683,202 and 4,965,188.

The identification of biallelic markers as described above allows the design of appropriate oligonucleotides, which can be used as primers to amplify DNA fragments comprising the biallelic markers of the present invention. Amplification can be performed using the primers initially used to discover new biallelic markers which are described herein or any set of primers allowing the
5 amplification of a DNA fragment comprising a biallelic marker of the present invention. Primers can be prepared by any suitable method. As for example, direct chemical synthesis by a method such as the phosphodiester method of Narang S.A. et al. (*Methods Enzymol.* 68:90-98, 1979), the phosphodiester method of Brown E.L. et al. (*Methods Enzymol.* 68:109-151, 1979), the diethylphosphoramidite method of Beaucage et al. (*Tetrahedron Lett.* 22:1859-1862, 1981) and the
10 solid support method described in EP 0 707 592.

In some embodiments the present invention provides primers for amplifying a DNA fragment containing one or more biallelic markers of the present invention. It will be appreciated that the amplification primers listed in the present specification are merely exemplary and that any other set of primers which produce amplification products containing one or more biallelic markers of the present
15 invention.

The primers are selected to be substantially complementary to the different strands of each specific sequence to be amplified. The length of the primers of the present invention can range from 8 to 100 nucleotides, preferably from 8 to 50, 8 to 30 or more preferably 8 to 25 nucleotides. Shorter primers tend to lack specificity for a target nucleic acid sequence and generally require cooler
20 temperatures to form sufficiently stable hybrid complexes with the template. Longer primers are expensive to produce and can sometimes self-hybridize to form hairpin structures. The formation of stable hybrids depends on the melting temperature T_m of the DNA. The T_m depends on the length of the primer, the ionic strength of the solution and the G+C content. The higher the G+C content of the primer, the higher is the melting temperature because G:C pairs are held by three H bonds whereas
25 A:T pairs have only two. The G+C content of the amplification primers of the present invention preferably ranges between 10 and 75 %, more preferably between 35 and 60 %, and most preferably between 40 and 55 %. The appropriate length for primers under a particular set of assay conditions is empirically determined by one of skill in the art.

The spacing of the primers determines the length of the segment to be amplified. In the
30 context of the present invention amplified segments carrying biallelic markers can range in size from at least about 25 bp to 35 kbp. Amplification fragments from 25-3000 bp are typical, fragments from 50-1000 bp are preferred and fragments from 100-600 bp are highly preferred. It will be appreciated that amplification primers for the biallelic markers is any sequence which allow the specific amplification of any DNA fragment carrying the markers. Amplification primers is labeled or
35 immobilized on a solid support as described in Section II.

X.C. Methods of Genotyping DNA samples for Biallelic Markers

Any method known in the art can be used to identify the nucleotide present at a biallelic marker site. Since the biallelic marker allele to be detected has been identified and specified in the present invention, detection will prove routine for one of ordinary skill in the art by employing any of a number of techniques. Many genotyping methods require the previous amplification of the DNA region carrying the biallelic marker of interest. While the amplification of target or signal is often preferred at present, ultrasensitive detection methods which do not require amplification are also encompassed by the present genotyping methods. Methods well-known to those skilled in the art that can be used to detect biallelic polymorphisms include methods such as, conventional dot blot analyzes, single strand conformational polymorphism analysis (SSCP) described by Orita et al. (*Proc. Natl. Acad. Sci. U.S.A* 86:27776-2770, 1989), denaturing gradient gel electrophoresis (DGGE), heteroduplex analysis, mismatch cleavage detection, and other conventional techniques as described in Sheffield, V.C. et al. (*Proc. Natl. Acad. Sci. USA* 49:699-706, 1991), White et al. (*Genomics* 12:301-306, 1992), Grompe, M. et al. (*Proc. Natl. Acad. Sci. USA* 86:5855-5892, 1989) and Grompe, M. (*Nature Genetics* 5:111-117, 1993). Another method for determining the identity of the nucleotide present at a particular polymorphic site employs a specialized exonuclease-resistant nucleotide derivative as described in US patent 4,656,127.

Preferred methods involve directly determining the identity of the nucleotide present at a biallelic marker site by sequencing assay, allele-specific amplification assay, or hybridization assay. The following is a description of some preferred methods. A highly preferred method is the microsequencing technique. The term "sequencing assay" is used herein to refer to polymerase extension of duplex primer/template complexes and includes both traditional sequencing and microsequencing.

1) Sequencing assays

The nucleotide present at a polymorphic site can be determined by sequencing methods. In a preferred embodiment, DNA samples are subjected to PCR amplification before sequencing as described above. Methods for sequencing DNA using either the dideoxy-mediated method (Sanger method) or the Maxam-Gilbert method are widely known to those of ordinary skill in the art. Such methods are for example disclosed in Maniatis et al. (*Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Second Edition, 1989). Alternative approaches include hybridization to high-density DNA probe arrays as described in Chee et al. (*Science* 274, 610, 1996).

Preferably, the amplified DNA is subjected to automated dideoxy terminator sequencing reactions using a dye-primer cycle sequencing protocol. The products of the sequencing reactions are run on sequencing gels and the sequences are determined using gel image analysis.

The polymorphism detection in a pooled sample is based on the presence of superimposed peaks in the electrophoresis pattern resulting from different bases occurring at the same position. Because each dideoxy terminator is labeled with a different fluorescent molecule, the two peaks

corresponding to a biallelic site present distinct colors corresponding to two different nucleotides at the same position on the sequence. However, the presence of two peaks can be an artifact due to background noise. To exclude such an artifact, the two DNA strands are sequenced and a comparison between the peaks is carried out. In order to be registered as a polymorphic sequence, the polymorphism has to be detected on both strands.

The above procedure permits those amplification products, which contain biallelic markers to be identified. The detection limit for the frequency of biallelic polymorphisms detected by sequencing pools of 100 individuals is approximately 0.1 for the minor allele, as verified by sequencing pools of known allelic frequencies.

10 Microsequencing assays

In microsequencing methods, the nucleotide at a polymorphic site in a target DNA is detected by a single nucleotide primer extension reaction. This method involves appropriate microsequencing primers which, hybridize just upstream of the polymorphic base of interest in the target nucleic acid. A polymerase is used to specifically extend the 3' end of the primer with one single ddNTP (chain terminator) complementary to the nucleotide at the polymorphic site. Next the identity of the incorporated nucleotide is determined in any suitable way.

Typically, microsequencing reactions are carried out using fluorescent ddNTPs and the extended microsequencing primers are analyzed by electrophoresis on ABI 377 sequencing machines to determine the identity of the incorporated nucleotide as described in EP 412 883. Alternatively, capillary electrophoresis can be used in order to process a higher number of assays simultaneously.

Different approaches can be used to detect the nucleotide added to the microsequencing primer. A homogeneous phase detection method based on fluorescence resonance energy transfer has been described by Chen and Kwok (*Nucleic Acids Research* 25:347-353 1997) and Chen et al. (*Proc. Natl. Acad. Sci. USA* 94/20 10756-10761, 1997). In this method amplified genomic DNA fragments containing polymorphic sites are incubated with a 5'-fluorescein-labeled primer in the presence of allelic dye-labeled dideoxynucleoside triphosphates and a modified Taq polymerase. The dye-labeled primer is extended one base by the dye-terminator specific for the allele present on the template. At the end of the genotyping reaction, the fluorescence intensities of the two dyes in the reaction mixture are analyzed directly without separation or purification. All these steps can be performed in the same tube and the fluorescence changes can be monitored in real time. Alternatively, the extended primer is analyzed by MALDI-TOF Mass Spectrometry. The base at the polymorphic site is identified by the mass added onto the microsequencing primer (see Haff L.A. and Smirnov I.P., *Genome Research*, 7:378-388, 1997).

Microsequencing is achieved by the established microsequencing method or by developments or derivatives thereof. Alternative methods include several solid-phase microsequencing techniques. The basic microsequencing protocol is the same as described previously, except that the method is

conducted as a heterogenous phase assay, in which the primer or the target molecule is immobilized or captured onto a solid support. To simplify the primer separation and the terminal nucleotide addition analysis, oligonucleotides are attached to solid supports or are modified in such ways that permit affinity separation as well as polymerase extension. The 5' ends and internal nucleotides of synthetic oligonucleotides can be modified in a number of different ways to permit different affinity separation approaches, e.g., biotinylation. If a single affinity group is used on the oligonucleotides, the oligonucleotides can be separated from the incorporated terminator reagent. This eliminates the need of physical or size separation. More than one oligonucleotide can be separated from the terminator reagent and analyzed simultaneously if more than one affinity group is used. This permits the analysis of several nucleic acid species or more nucleic acid sequence information per extension reaction. The affinity group need not be on the priming oligonucleotide but could alternatively be present on the template. For example, immobilization can be carried out via an interaction between biotinylated DNA and streptavidin-coated microtitration wells or avidin-coated polystyrene particles. In the same manner oligonucleotides or templates is attached to a solid support in a high-density format. In such solid phase microsequencing reactions, incorporated ddNTPs can be radiolabeled (Syvänen, *Clinica Chimica Acta* 226:225-236, 1994) or linked to fluorescein (Livak and Hainer, *Human Mutation* 3:379-385, 1994). The detection of radiolabeled ddNTPs can be achieved through scintillation-based techniques. The detection of fluorescein-linked ddNTPs can be based on the binding of anti fluorescein antibody conjugated with alkaline phosphatase, followed by incubation with a chromogenic substrate (such as *p*-nitrophenyl phosphate). Other possible reporter-detection pairs include: ddNTP linked to dinitrophenyl (DNP) and anti-DNP alkaline phosphatase conjugate (Harju et al., *Clin. Chem.* 39/11 2282-2287, 1993) or biotinylated ddNTP and horseradish peroxidase-conjugated streptavidin with *o*-phenylenediamine as a substrate (WO 92/15712). As yet another alternative solid-phase microsequencing procedure, Nyren et al. (*Analytical Biochemistry* 208:171-175, 1993) described a method relying on the detection of DNA polymerase activity by an enzymatic luminometric inorganic pyrophosphate detection assay (ELIDA).

Pastinen et al. (*Genome research* 7:606-614, 1997) describe a method for multiplex detection of single nucleotide polymorphism in which the solid phase minisequencing principle is applied to an oligonucleotide array format. High-density arrays of DNA probes attached to a solid support (DNA chips) are further described in X.C.5.

In one aspect the present invention provides polynucleotides and methods to genotype one or more biallelic markers of the present invention by performing a microsequencing assay. It will be appreciated that any primer having a 3' end immediately adjacent to the polymorphic nucleotide is used. However, polynucleotides comprising at least 8, 12, 15, 20, 25, or 30 consecutive nucleotides of the sequence immediately adjacent to the biallelic marker and having a 3' terminus immediately

upstream of the corresponding biallelic marker are well suited for determining the identity of a nucleotide at biallelic marker site.

Similarly, it will be appreciated that microsequencing analysis is performed for any biallelic marker or any combination of biallelic markers of the present invention.

5 Mismatch detection assays based on polymerases and ligases

In one aspect the present invention provides polynucleotides and methods to determine the allele of one or more biallelic markers of the present invention in a biological sample, by mismatch detection assays based on polymerases and/or ligases. These assays are based on the specificity of polymerases and ligases. Polymerization reactions places particularly stringent requirements on
10 correct base pairing of the 3' end of the amplification primer and the joining of two oligonucleotides hybridized to a target DNA sequence is quite sensitive to mismatches close to the ligation site, especially at the 3' end. Methods, primers and various parameters to amplify DNA fragments comprising biallelic markers of the present invention are further described above in X.B.

Allele specific amplification

15 Discrimination between the two alleles of a biallelic marker can also be achieved by allele specific amplification, a selective strategy, whereby one of the alleles is amplified without amplification of the other allele. This is accomplished by placing the polymorphic base at the 3' end of one of the amplification primers. Because the extension forms from the 3' end of the primer, a mismatch at or near this position has an inhibitory effect on amplification. Therefore, under
20 appropriate amplification conditions, these primers only direct amplification on their complementary allele. Designing the appropriate allele-specific primer and the corresponding assay conditions are well with the ordinary skill in the art.

Ligation/amplification based methods

The "Oligonucleotide Ligation Assay" (OLA) uses two oligonucleotides which are designed
25 to be capable of hybridizing to abutting sequences of a single strand of a target molecules. One of the oligonucleotides is biotinylated, and the other is detectably labeled. If the precise complementary sequence is found in a target molecule, the oligonucleotides will hybridize such that their termini abut, and create a ligation substrate that can be captured and detected. OLA is capable of detecting single nucleotide polymorphisms and is advantageously combined with PCR as described by
30 Nickerson D.A. et al. (*Proc. Natl. Acad. Sci. U.S.A.* 87:8923-8927, 1990). In this method, PCR is used to achieve the exponential amplification of target DNA, which is then detected using OLA.

Other methods which are particularly suited for the detection of single nucleotide polymorphism include LCR (ligase chain reaction), Gap LCR (GLCR) which are described above in X.B. As mentioned above LCR uses two pairs of probes to exponentially amplify a specific target.
35 The sequences of each pair of oligonucleotides, is selected to permit the pair to hybridize to abutting sequences of the same strand of the target. Such hybridization forms a substrate for a template-

dependant ligase. In accordance with the present invention, LCR can be performed with oligonucleotides having the proximal and distal sequences of the same strand of a biallelic marker site. In one embodiment, either oligonucleotide will be designed to include the biallelic marker site. In such an embodiment, the reaction conditions are selected such that the oligonucleotides can be
5 ligated together only if the target molecule either contains or lacks the specific nucleotide that is complementary to the biallelic marker on the oligonucleotide. In an alternative embodiment, the oligonucleotides will not include the biallelic marker, such that when they hybridize to the target molecule, a "gap" is created as described in WO 90/01069. This gap is then "filled" with complementary dNTPs (as mediated by DNA polymerase), or by an additional pair of
10 oligonucleotides. Thus at the end of each cycle, each single strand has a complement capable of serving as a target during the next cycle and exponential allele-specific amplification of the desired sequence is obtained.

Ligase/Polymerase-mediated Genetic Bit AnalysisTM is another method for determining the identity of a nucleotide at a preselected site in a nucleic acid molecule (WO 95/21271). This method
15 involves the incorporation of a nucleoside triphosphate that is complementary to the nucleotide present at the preselected site onto the terminus of a primer molecule, and their subsequent ligation to a second oligonucleotide. The reaction is monitored by detecting a specific label attached to the reaction's solid phase or by detection in solution.

2) Hybridization assay methods

20 A preferred method of determining the identity of the nucleotide present at a biallelic marker site involves nucleic acid hybridization. The hybridization probes, which can be conveniently used in such reactions, preferably include the probes defined herein. Any hybridization assay is used including Southern hybridization, Northern hybridization, dot blot hybridization and solid-phase hybridization (see Sambrook et al., Molecular Cloning – A Laboratory Manual, Second Edition, Cold
25 Spring Harbor Press, N.Y., 1989).

Hybridization refers to the formation of a duplex structure by two single stranded nucleic acids due to complementary base pairing. Hybridization can occur between exactly complementary nucleic acid strands or between nucleic acid strands that contain minor regions of mismatch. Specific probes can be designed that hybridize to one form of a biallelic marker and not to the other and
30 therefore are able to discriminate between different allelic forms. Allele-specific probes are often used in pairs, one member of a pair showing perfect match to a target sequence containing the original allele and the other showing a perfect match to the target sequence containing the alternative allele. Hybridization conditions should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a
35 probe hybridizes to only one of the alleles. Stringent, sequence specific hybridization conditions, under which a probe will hybridize only to the exactly complementary target sequence are well known

in the art (Sambrook et al., *Molecular Cloning - A Laboratory Manual*, Second Edition, Cold Spring Harbor Press, N.Y., 1989). Stringent conditions are sequence dependent and will be different in different circumstances. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point TM for the specific sequence at a defined ionic strength and pH. By way of example and not limitation, procedures using conditions of high stringency are as follows:

Prehybridization of filters containing DNA is carried out for 8 h to overnight at 65°C in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 µg/ml denatured salmon sperm DNA. Filters are hybridized for 48 h at 65°C, the preferred hybridization temperature, in prehybridization mixture containing 100 µg/ml denatured salmon sperm DNA and 5-20 X 10⁶ cpm of ³²P-labeled probe. Alternatively, the hybridization step can be performed at 65°C in the presence of SSC buffer, 1 x SSC corresponding to 0.15M NaCl and 0.05 M Na citrate. Subsequently, filter washes can be done at 37°C for 1 h in a solution containing 2X SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA, followed by a wash in 0.1X SSC at 50°C for 45 min. Alternatively, filter washes can be performed in a solution containing 2 x SSC and 0.1% SDS, or 0.5 x SSC and 0.1% SDS, or 0.1 x SSC and 0.1% SDS at 68°C for 15 minute intervals. Following the wash steps, the hybridized probes are detectable by autoradiography. By way of example and not limitation, procedures using conditions of intermediate stringency are as follows: Filters containing DNA are prehybridized, and then hybridized at a temperature of 60°C in the presence of a 5 x SSC buffer and labeled probe. Subsequently, filters washes are performed in a solution containing 2x SSC at 50°C and the hybridized probes are detectable by autoradiography. Other conditions of high and intermediate stringency which is used are well known in the art and as cited in Sambrook et al. (*Molecular Cloning - A Laboratory Manual*, Second Edition, Cold Spring Harbor Press, N.Y., 1989) and Ausubel et al. (*Current Protocols in Molecular Biology*, Green Publishing Associates and Wiley Interscience, N.Y., 1989).

Although such hybridizations can be performed in solution, it is preferred to employ a solid-phase hybridization assay. The target DNA comprising a biallelic marker of the present invention is amplified prior to the hybridization reaction. The presence of a specific allele in the sample is determined by detecting the presence or the absence of stable hybrid duplexes formed between the probe and the target DNA. The detection of hybrid duplexes can be carried out by a number of methods. Various detection assay formats are well known which utilize detectable labels bound to either the target or the probe to enable detection of the hybrid duplexes. Typically, hybridization duplexes are separated from unhybridized nucleic acids and the labels bound to the duplexes are then detected. Those skilled in the art will recognize that wash steps is employed to wash away excess target DNA or probe. Standard heterogeneous assay formats are suitable for detecting the hybrids using the labels present on the primers and probes.

Two recently developed assays allow hybridization-based allele discrimination with no need for separations or washes (see Landegren U. et al., *Genome Research*, 8:769-776,1998). The TaqMan assay takes advantage of the 5' nuclease activity of Taq DNA polymerase to digest a DNA probe annealed specifically to the accumulating amplification product. TaqMan probes are labeled with a donor-acceptor dye pair that interacts via fluorescence energy transfer. Cleavage of the TaqMan probe by the advancing polymerase during amplification dissociates the donor dye from the quenching acceptor dye, greatly increasing the donor fluorescence. All reagents necessary to detect two allelic variants can be assembled at the beginning of the reaction and the results are monitored in real time (see Livak et al., *Nature Genetics*, 9:341-342, 1995). In an alternative homogeneous hybridization based procedure, molecular beacons are used for allele discriminations. Molecular beacons are hairpin-shaped oligonucleotide probes that report the presence of specific nucleic acids in homogeneous solutions. When they bind to their targets they undergo a conformational reorganization that restores the fluorescence of an internally quenched fluorophore (Tyagi et al., *Nature Biotechnology*, 16:49-53, 1998).

The polynucleotides provided herein can be used in hybridization assays for the detection of biallelic marker alleles in biological samples. These probes are characterized in that they preferably comprise between 8 and 50 nucleotides, and in that they are sufficiently complementary to a sequence comprising a biallelic marker of the present invention to hybridize thereto and preferably sufficiently specific to be able to discriminate the targeted sequence for only one nucleotide variation. The GC content in the probes of the invention usually ranges between 10 and 75 %, preferably between 35 and 60 %, and more preferably between 40 and 55 %. The length of these probes can range from 10, 15, 20, or 30 to at least 100 nucleotides, preferably from 10 to 50, more preferably from 18 to 35 nucleotides. A particularly preferred probe is 25 nucleotides in length. Preferably the biallelic marker is within 4 nucleotides of the center of the polynucleotide probe. In particularly preferred probes the biallelic marker is at the center of said polynucleotide. Shorter probes may lack specificity for a target nucleic acid sequence and generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. Longer probes are expensive to produce and can sometimes self-hybridize to form hairpin structures. Methods for the synthesis of oligonucleotide probes have been described above and can be applied to the probes of the present invention.

Preferably the probes of the present invention are labeled or immobilized on a solid support. Labels and solid supports are further described in II. Detection probes are generally nucleic acid sequences or uncharged nucleic acid analogs such as, for example peptide nucleic acids which are disclosed in International Patent Application WO 92/20702, morpholino analogs which are described in U.S. Patents Numbered 5,185,444; 5,034,506 and 5,142,047. The probe may have to be rendered "non-extendable" in that additional dNTPs cannot be added to the probe. In and of themselves analogs usually are non-extendable and nucleic acid probes can be rendered non-extendable by

modifying the 3' end of the probe such that the hydroxyl group is no longer capable of participating in elongation. For example, the 3' end of the probe can be functionalized with the capture or detection label to thereby consume or otherwise block the hydroxyl group. Alternatively, the 3' hydroxyl group simply can be cleaved, replaced or modified, U.S. Patent Application Serial No. 07/049,061 filed 5 April 19, 1993 describes modifications, which can be used to render a probe non-extendable.

The probes of the present invention are useful for a number of purposes. They can be used in Southern hybridization to genomic DNA or Northern hybridization to mRNA. The probes can also be used to detect PCR amplification products. By assaying the hybridization to an allele specific probe, one can detect the presence or absence of a biallelic marker allele in a given sample.

10 High-Throughput parallel hybridizations in array format are specifically encompassed within "hybridization assays" and are described below.

Hybridization to addressable arrays of oligonucleotides

Hybridization assays based on oligonucleotide arrays rely on the differences in hybridization stability of short oligonucleotides to perfectly matched and mismatched target sequence variants. 15 Efficient access to polymorphism information is obtained through a basic structure comprising high-density arrays of oligonucleotide probes attached to a solid support (the chip) at selected positions. Each DNA chip can contain thousands to millions of individual synthetic DNA probes arranged in a grid-like pattern and miniaturized to the size of a dime.

The chip technology has already been applied with success in numerous cases. For example, 20 the screening of mutations has been undertaken in the BRCA1 gene, in *S. cerevisiae* mutant strains, and in the protease gene of HIV-1 virus (Hacia et al., *Nature Genetics*, 14(4):441-447, 1996; Shoemaker et al., *Nature Genetics*, 14(4):450-456, 1996; Kozal et al., *Nature Medicine*, 2:753-759, 1996). Chips of various formats for use in detecting biallelic polymorphisms can be produced on a customized basis by Affymetrix (GeneChip™), Hyseq (HyChip and HyGnostics), and Protogene 25 Laboratories.

In general, these methods employ arrays of oligonucleotide probes that are complementary to target nucleic acid sequence segments from an individual which, target sequences include a polymorphic marker. EP785280 describes a tiling strategy for the detection of single nucleotide polymorphisms. Briefly, arrays may generally be "tiled" for a large number of specific 30 polymorphisms. By "tiling" is generally meant the synthesis of a defined set of oligonucleotide probes which is made up of a sequence complementary to the target sequence of interest, as well as preselected variations of that sequence, e.g., substitution of one or more given positions with one or more members of the basis set of monomers, i.e. nucleotides. Tiling strategies are further described in PCT application No. WO 95/11995. In a particular aspect, arrays are tiled for a number of specific, 35 identified biallelic marker sequences. In particular the array is tiled to include a number of detection blocks, each detection block being specific for a specific biallelic marker or a set of biallelic markers.

For example, a detection block is tiled to include a number of probes, which span the sequence segment that includes a specific polymorphism. To ensure probes that are complementary to each allele, the probes are synthesized in pairs differing at the biallelic marker. In addition to the probes differing at the polymorphic base, monosubstituted probes are also generally tiled within the detection block. These monosubstituted probes have bases at and up to a certain number of bases in either direction from the polymorphism, substituted with the remaining nucleotides (selected from A, T, G, C and U). Typically the probes in a tiled detection block will include substitutions of the sequence positions up to and including those that are 5 bases away from the biallelic marker. The monosubstituted probes provide internal controls for the tiled array, to distinguish actual hybridization from artefactual cross-hybridization. Upon completion of hybridization with the target sequence and washing of the array, the array is scanned to determine the position on the array to which the target sequence hybridizes. The hybridization data from the scanned array is then analyzed to identify which allele or alleles of the biallelic marker are present in the sample. Hybridization and scanning is carried out as described in PCT application No. WO 92/10092 and WO 95/11995 and US patent No. 5,424,186.

5) Integrated Systems

Another technique, which is used to analyze polymorphisms, includes multicomponent integrated systems, which miniaturize and compartmentalize processes such as PCR and capillary electrophoresis reactions in a single functional device. An example of such technique is disclosed in US patent 5,589,136, which describes the integration of PCR amplification and capillary electrophoresis in chips.

Integrated systems can be envisaged mainly when microfluidic systems are used. These systems comprise a pattern of microchannels designed onto a glass, silicon, quartz, or plastic wafer included on a microchip. The movements of the samples are controlled by electric, electroosmotic or hydrostatic forces applied across different areas of the microchip to create functional microscopic valves and pumps with no moving parts. Varying the voltage controls the liquid flow at intersections between the micro-machined channels and changes the liquid flow rate for pumping across different sections of the microchip.

For genotyping biallelic markers, the microfluidic system may integrate nucleic acid amplification, microsequencing, capillary electrophoresis and a detection method such as laser-induced fluorescence detection.

XI. METHODS OF GENETIC ANALYSIS USING THE BIALLELIC MARKERS OF THE PRESENT INVENTION

The methods available for the genetic analysis of complex traits fall into different categories (see Lander and Schork, *Science*, 265, 2037-2048, 1994). In general, the biallelic markers of the present invention find use in any method known in the art to demonstrate a statistically significant

correlation between a genotype and a phenotype. The biallelic markers is used in linkage analysis and in allele-sharing methods. Preferably, the biallelic markers of the present invention are used to identify genes associated with detectable traits using association studies, an approach which does not require the use of affected families and which permits the identification of genes associated with
5 complex and sporadic traits.

The genetic analysis using the biallelic markers of the present invention is conducted on any scale. The whole set of biallelic markers of the present invention or any subset of biallelic markers of the present invention is used. In some embodiments, any additional set of genetic markers including a biallelic marker of the present invention is used. As mentioned above, it should be noted that the
10 biallelic markers of the present invention is included in any complete or partial genetic map of the human genome. These different uses are specifically contemplated in the present invention and claims.

XI.A. Linkage Analysis

Until recently, the identification of genes linked with detectable traits has mainly relied on a
15 statistical approach called linkage analysis. Linkage analysis involves proposing a model to explain the inheritance pattern of phenotypes and genotypes observed in a pedigree. Linkage analysis is based upon establishing a correlation between the transmission of genetic markers and that of a specific trait throughout generations within a family. In this approach, all members of a series of affected families are genotyped with a few hundred markers, typically microsatellite markers, which
20 are distributed at an average density of one every 10 Mb. By comparing genotypes in all family members, one can attribute sets of alleles to parental haploid genomes (haplotyping or phase determination). The origin of recombined fragments is then determined in the offspring of all families. Those that co-segregate with the trait are tracked. After pooling data from all families, statistical methods are used to determine the likelihood that the marker and the trait are segregating
25 independently in all families. As a result of the statistical analysis, one or several regions having a high probability of harboring a gene linked to the trait are selected as candidates for further analysis. The result of linkage analysis is considered as significant (i.e. there is a high probability that the region contains a gene involved in a detectable trait) when the chance of independent segregation of the marker and the trait is lower than 1 in 1000 (expressed as a LOD score > 3). Generally, the length
30 of the candidate region identified as having a LOD score of greater than 3 using linkage analysis is between 2 and 20Mb. Once a candidate region is identified as described above, analysis of recombinant individuals using additional markers allows further delineation of the candidate region. Linkage analysis studies have generally relied on the use of a maximum of 5,000 microsatellite markers, thus limiting the maximum theoretical attainable resolution of linkage analysis to about 600
35 kb on average.

Linkage analysis has been successfully applied to map simple genetic traits that show clear Mendelian inheritance patterns and which have a high penetrance (i.e., the ratio between the number of trait positive carriers of allele a and the total number of a carriers in the population). About 100 pathological trait-causing genes were discovered using linkage analysis over the last 10 years. In most of these cases, the majority of affected individuals had affected relatives and the detectable trait was rare in the general population (frequencies less than 0.1%). In about 10 cases, such as Alzheimer's Disease, breast cancer, and Type II diabetes, the detectable trait was more common but the allele associated with the detectable trait was rare in the affected population. Thus, the alleles associated with these traits were not responsible for the trait in all sporadic cases.

Linkage analysis suffers from a variety of drawbacks. First, linkage analysis is limited by its reliance on the choice of a genetic model suitable for each studied trait. Furthermore, as already mentioned, the resolution attainable using linkage analysis is limited, and complementary studies are required to refine the analysis of the typical 2Mb to 20Mb regions initially identified through linkage analysis. In addition, linkage analysis approaches have proven difficult when applied to complex genetic traits, such as those due to the combined action of multiple genes and/or environmental factors. In such cases, too large an effort and cost are needed to recruit the adequate number of affected families required for applying linkage analysis to these situations, as recently discussed by Risch, N. and Merikangas, K. (*Science*, 273:1516-1517, 1996). Finally, linkage analysis cannot be applied to the study of traits for which no large informative families are available. Typically, this will be the case in any attempt to identify trait-causing alleles involved in sporadic cases, such as alleles associated with positive or negative responses to drug treatment.

XI.B. Allele-Sharing methods

Whereas linkage analysis involves proposing a model to explain the inheritance pattern of phenotypes and genotypes in a pedigree, allele-sharing methods are not based on constructing a model, but rather on rejecting a model (see Lander and Schork, *Science*, 265, 2037-2048, 1994). More specifically, one tries to prove that the inheritance pattern of a chromosomal region is not consistent with random Mendelian segregation by showing that affected relatives inherit identical copies of the region more often than expected by chance. Because allele-sharing methods are nonparametric (that is, assume no model for the inheritance of the trait), they tend to be more useful for the analysis of complex traits than linkage analysis. Affected relatives should show excess allele sharing even in the presence of incomplete penetrance and polygenic inheritance. Allele-Sharing methods involve studying affected relatives in a pedigree to determine how often a particular copy of a chromosomal region is shared identical-by-descent (IBD), that is, is inherited from a common ancestor within the pedigree. The frequency of IBD sharing at a locus can then be compared with random expectation. Affected sib pair analysis is a well-known special case and is the simplest form of this method.

However, as allele-sharing methods analyze affected relatives, they tend to be of limited value in the genetic analysis of drug responses or in the analysis of side effects to treatments. This type of analysis is impractical in such cases due to the lack of availability of familial cases. In fact, the likelihood of having more than one individual in a family being exposed to the same drug at the same time is very low.

XI.C. Association Studies

The present invention comprises methods for identifying one or several genes among a set of candidate genes that are associated with a detectable trait using the biallelic markers of the present invention. In one embodiment the present invention comprises methods to detect an association between a biallelic marker allele or a biallelic marker haplotype and a trait. Further, the invention comprises methods to identify a trait causing allele in linkage disequilibrium with any biallelic marker allele of the present invention.

As described above, alternative approaches can be employed to perform association studies: genome-wide association studies, candidate region association studies and candidate gene association studies. In a preferred embodiment, the biallelic markers of the present invention are used to perform candidate gene association studies. The candidate gene analysis clearly provides a short-cut approach to the identification of genes and gene polymorphisms related to a particular trait when some information concerning the biology of the trait is available. Further, the biallelic markers of the present invention is incorporated in any map of genetic markers of the human genome in order to perform genome-wide association studies. Methods to generate a high-density map of biallelic markers has been described in US Provisional Patent application serial number 60/082,614. The biallelic markers of the present invention may further be incorporated in any map of a specific candidate region of the genome (a specific chromosome or a specific chromosomal region for example).

As mentioned above, association studies is conducted within the general population and are not limited to studies performed on related individuals in affected families. Linkage disequilibrium and association studies are extremely valuable as they permit the analysis of sporadic or multifactor traits. Moreover, association studies represent a powerful method for fine-scale mapping enabling much finer mapping of trait causing alleles than linkage studies. Studies based on pedigrees often only narrow the location of the trait causing allele. Association studies and Linkage Disequilibrium mapping methods using the biallelic markers of the present invention can therefore be used to refine the location of a trait causing allele in a candidate region identified by Linkage Analysis or by Allele-Sharing methods. Moreover, once a chromosome segment of interest has been identified, the presence of a candidate gene such as a candidate gene of the present invention, in the region of interest can provide a shortcut to the identification of the trait causing allele. Biallelic markers of the

present invention can be used to demonstrate that a candidate gene is associated with a trait. Such uses are specifically contemplated in the present invention and claims.

1) Case-control populations (inclusion criteria)

Association studies do not concern familial inheritance and do not involve the analysis of large family pedigrees but compare the prevalence of a particular genetic marker, or a set of markers, in case-control populations. They are case-control studies based on comparison of unrelated case (affected or trait positive) individuals and unrelated control (random or unaffected or trait negative) individuals. The control group is composed of individuals chosen randomly or of unaffected (trait negative) individuals, preferably the control group is composed of unaffected or trait negative individuals. Further, the control group is preferably both ethnically- and age-matched to the case population. In the following "trait positive population", "case population" and "affected population" are used interchangeably.

An important step in the dissection of complex traits using association studies is the choice of case-control populations (see Lander and Schork, *Science*, 265, 2037-2048, 1994). Narrowing the definition of the disease and restricting the patient population to extreme phenotypes allows one to work with a trait that is more nearly Mendelian in its inheritance pattern and more likely to be homogeneous (patients suffer from the disease for the same genetic reasons). Therefore, a major step in the choice of case-control populations is the clinical definition of a given trait or phenotype. Four criteria are often useful: clinical phenotype, age at onset, family history and severity. Preferably, in order to perform efficient and significant association studies, such as those described herein, the trait under study should preferably follow a bimodal distribution in the population under study, presenting two clear non-overlapping phenotypes (trait positive and trait negative). Nevertheless, even in the absence of such bimodal distribution (as may in fact be the case for more complex genetic traits), any genetic trait may still be analyzed by the association method proposed here by carefully selecting the individuals to be included in the trait positive and trait negative phenotypic groups. The selection procedure involves selecting individuals at opposite ends of the non-bimodal phenotype spectra of the trait under study, so as to include in these trait positive and trait negative populations individuals which clearly represent extreme, preferably non-overlapping phenotypes. This is particularly useful for continuous or quantitative traits (such as blood pressure for example). Selection of individuals at extreme ends of the trait distribution increases the ability to analyze these complex traits. The definition of the inclusion criteria for the case-control populations is an important aspect of association studies. The selection of those drastically different but relatively uniform phenotypes enables efficient comparisons in association studies and the possible detection of marked differences at the genetic level, provided that the sample sizes of the populations under study are significant enough.

Preferably, case-control populations to be included in association studies such as those proposed in the present invention consist of phenotypically homogeneous populations of individuals each representing 100% of the corresponding phenotype if the trait distribution is bimodal. If the trait distribution is non-bimodal, trait positive and trait negative populations consist of phenotypically uniform populations of individuals representing each between 1 and 98%, preferably between 1 and 80%, more preferably between 1 and 50%, and more preferably between 1 and 30%, most preferably between 1 and 20% of the total population under study, and selected among individuals exhibiting non-overlapping phenotypes. In some embodiments, the trait positive and trait negative groups consist of individuals exhibiting the extreme phenotypes within the studied population. The clearer the difference between the two trait phenotypes, the greater the probability of detecting an association with biallelic markers.

In preferred embodiments, a first group of between 50 and 300 trait positive individuals, preferably about 100 individuals, are recruited according to their phenotypes. A similar number of trait negative individuals are included in such studies.

In the present invention, typical examples of inclusion criteria include a diagnosis of cancer or prostate cancer or the evaluation of the response to anti-cancer or anti-prostate cancer agent or side effects to treatment with anti-cancer or anti-prostate cancer agents.

Suitable examples of association studies using biallelic markers including the biallelic markers of the present invention, are studies involving the following populations:

- a case population suffering from a form of cancer and a healthy unaffected control population, or
- a case population suffering from a form of prostate cancer and a healthy unaffected control population, or
- a case population treated with anticancer agents suffering from side-effects resulting from the treatment and a control population treated with the same agents showing no side-effects, or
- a case population treated with anti-prostate cancer agents suffering from side-effects resulting from the treatment and a control population treated with the same agents showing no side-effects, or
- a case population treated with anti-cancer agents showing a beneficial response and a control population treated with same agents showing no beneficial response, or
- a case population treated with anti-prostate cancer agents showing a beneficial response and a control population treated with same agents showing no beneficial response.

2) Determining the frequency of an allele in case-control populations

Allelic frequencies of the biallelic markers in each of the populations can be determined using one of the methods described above under the in Section X. under the heading "Methods for genotyping an individual for biallelic markers", or any genotyping procedure suitable for this intended purpose. The frequency of a biallelic marker allele in a population can be determined by genotyping pooled samples or individual samples. One way to reduce the number of genotypings required is to

use pooled samples. A major obstacle in using pooled samples is in terms of accuracy and reproducibility for determining accurate DNA concentrations in setting up the pools. Genotyping individual samples provides higher sensitivity, reproducibility and accuracy and; is the preferred method used in the present invention. Preferably, each individual is genotyped separately and simple
5 gene counting is applied to determine the frequency of an allele of a biallelic marker or of a genotype in a given population.

3) Determining the frequency of a haplotype in case-control populations

The gametic phase of haplotypes is usually unknown when diploid individuals are heterozygous at more than one locus. Different strategies for inferring haplotypes is used to partially
10 overcome this difficulty (see Excoffier L. and Slatkin M., *Mol. Biol. Evol.*, 12(5): 921-927, 1995). One possibility is that the multiple-site heterozygous diploids can be eliminated from the analysis, keeping only the homozygotes and the single-site heterozygote individuals, but this approach might lead to a possible bias in the sample composition and the underestimation of low-frequency haplotypes. Another possibility is that single chromosomes can be studied independently, for
15 example, by asymmetric PCR amplification (see Newton et al., *Nucleic Acids Res.*, 17:2503-2516, 1989; Wu et al., *Proc. Natl. Acad. Sci. USA*, 86:2757, 1989) or by isolation of single chromosome by limit dilution followed by PCR amplification (see Ruano et al., *Proc. Natl. Acad. Sci. USA*, 87:6296-6300, 1990). Further, multiple haplotypes can sometimes be inferred using genealogical information in families (Perlin et al., *Am. J. Hum. Genet.*, 55:777-787, 1994). A sample is haplotyped for
20 sufficiently close biallelic markers by double PCR amplification of specific alleles (Sarkar, G. and Sommer S.S., *Biotechniques*, 1991). These approaches are not entirely satisfying either because of their technical complexity, the additional cost they entail, their lack of generalization at a large scale, or the possible biases they introduce. To overcome these difficulties, an algorithm based on Hardy-Weinberg equilibrium (random mating) to infer the phase of PCR-amplified DNA genotypes
25 introduced by Clark A.G. (*Mol. Biol. Evol.*, 7:111-122, 1990) is used. Briefly, the principle is to start filling a preliminary list of haplotypes present in the sample by examining unambiguous individuals, that is, the complete homozygotes and the single-site heterozygotes. Then other individuals in the same sample are screened for the possible occurrence of previously recognized haplotypes. For each positive identification, the complementary haplotype is added to the list of recognized haplotypes,
30 until the phase information for all individuals is either resolved or identified as unresolved. This method assigns a single haplotype to each multiheterozygous individual, whereas several haplotypes are possible when there are more than one heterozygous site. Any other method known in the art to determine the frequency of a haplotype in a population is used. Preferably, an expectation-maximization (EM) algorithm (Dempster et al., *J. R. Stat. Soc.*, 39B:1-38, 1977) leading to maximum-likelihood estimates of haplotype frequencies under the assumption of Hardy-Weinberg proportions is
35 used (see Excoffier L. and Slatkin M., *Mol. Biol. Evol.*, 12(5): 921-927, 1995). The EM algorithm is

used to estimate haplotype frequencies in the case when only genotype data from unrelated individuals are available. The EM algorithm is a generalized iterative maximum-likelihood approach to estimation that is useful when data are ambiguous and/or incomplete. The EM algorithm is used to resolve heterozygotes into haplotypes. Haplotype estimations are further described below under the heading "Statistical methods".

4) Genetic Analysis based on Linkage Disequilibrium

Linkage disequilibrium is the non-random association of alleles at two or more loci and represents a powerful tool for genetic mapping of complex traits (see Jorde L.B., *Am. J. Hum. Genet.*, 56:11-14, 1995). Biallelic markers, because they are densely spaced in the human genome and can be genotyped in large numbers, are particularly useful in genetic analysis based on linkage disequilibrium.

When a disease mutation is first introduced into a population (by a new mutation or the immigration of a mutation carrier), it necessarily resides on a single chromosome and thus on a single "background" or "ancestral" haplotype of linked markers. Consequently, there is complete disequilibrium between these markers and the disease mutation: one finds the disease mutation only in the presence of a specific set of marker alleles. Through subsequent generations recombinations occur between the disease mutation and these marker polymorphisms, and the disequilibrium gradually dissipates. The pace of this dissipation is a function of the recombination frequency, so the markers closest to the disease gene will manifest higher levels of disequilibrium than those that are further away. When not broken up by recombination, "ancestral" haplotypes and linkage disequilibrium between marker alleles at different loci can be tracked not only through pedigrees but also through populations.

The pattern or curve of disequilibrium between disease and marker loci will exhibit a single maximum that occurs at the disease locus. Consequently, the amount of linkage disequilibrium between a disease allele and closely linked genetic markers may yield valuable information regarding the location of the disease gene. For fine-scale mapping of a disease locus, it is useful to have some knowledge of the patterns of linkage disequilibrium that exist between markers in the studied region. As mentioned above the mapping resolution achieved through the analysis of linkage disequilibrium is much higher than that of linkage studies. The high density of biallelic markers combined with linkage disequilibrium analysis provide powerful tools for fine-scale mapping. Different methods to calculate linkage disequilibrium are described below under the heading "Statistical Methods". Moreover, association studies as a method of mapping genetic traits rely on the phenomenon of linkage disequilibrium.

3) Association studies

As mentioned above, the occurrence of pairs of specific alleles at different loci on the same chromosome is not random, and the deviation from random is called linkage disequilibrium. If a

specific allele in a given gene is directly involved in causing a particular trait, its frequency will be statistically increased in an affected (trait positive) population when compared to the frequency in a trait negative population or in a random control population. As a consequence of the existence of linkage disequilibrium, the frequency of all other alleles present in the haplotype carrying the trait-causing allele will also be increased in trait positive individuals compared to trait negative individuals or random controls. Therefore, association between the trait and any allele (specifically a biallelic marker allele) in linkage disequilibrium with the trait-causing allele will suffice to suggest the presence of a trait-related gene in that particular allele's region. Association studies focus on population frequencies. Case-control populations can be genotyped for biallelic markers to identify associations that narrowly locate a trait causing allele. Moreover, any marker in linkage disequilibrium with one given marker associated with a trait will be associated with the trait. Linkage disequilibrium allows the relative frequencies in case-control populations of a limited number of genetic polymorphisms (specifically biallelic markers) to be analyzed as an alternative to screening all possible functional polymorphisms in order to find trait-causing alleles. Association studies compare the frequency of marker alleles in unrelated case-control populations, and represent powerful tools for the dissection of complex traits.

Association analysis

The general strategy to perform association studies using biallelic markers derived from a candidate gene is to scan two groups of individuals (case-control populations) in order to measure and statistically compare the allele frequencies of the biallelic markers of the present invention in both groups.

If a statistically significant association with a trait is identified for at least one or more of the analyzed biallelic markers, one can assume that: either the associated allele is directly responsible for causing the trait (the associated allele is the trait causing allele), or more likely the associated allele is in linkage disequilibrium with the trait causing allele. The specific characteristics of the associated allele with respect to the candidate gene function usually gives further insight into the relationship between the associated allele and the trait (causal or in linkage disequilibrium). If the evidence indicates that the associated allele within the candidate gene is most probably not the trait causing allele but is in linkage disequilibrium with the real trait causing allele, then the trait causing allele can be found by sequencing the vicinity of the associated marker.

Association studies are usually run in two successive steps. In a first phase, the frequencies of a reduced number of biallelic markers from one or several candidate genes are determined in the trait positive and trait negative populations. In a second phase of the analysis, the identity of the candidate gene and the position of the genetic loci responsible for the given trait is further refined using a higher density of markers from the relevant gene. However, if the candidate gene under study is relatively

small in length, as it is the case for many of the candidate genes analyzed included in the present invention, a single phase is sufficient to establish significant associations.

Haplotype analysis

As described above, when a chromosome carrying a disease allele first appears in a population as a result of either mutation or migration, the mutant allele necessarily resides on a chromosome having a unique set of linked markers: the ancestral haplotype. This haplotype can be tracked through populations and its statistical association with a given trait can be analyzed. The statistical power of association studies is increased by complementing single point (allelic) association studies with multi-point association studies also called haplotype studies. Thus, a haplotype association study allows one to define the frequency and the type of the ancestral carrier haplotype. A haplotype analysis is important in that it increases the statistical significance of an analysis involving individual markers. Indeed, by performing an association study with a set of biallelic markers, it increases the value of the results obtained through the study, allowing false positive and/or negative data that may result from the single marker studies to be eliminated.

In a first stage of a haplotype frequency analysis, the frequency of the possible haplotypes based on various combinations of the identified biallelic markers of the invention is determined. The haplotype frequency is then compared for distinct populations of trait positive and control individuals. The number of trait positive individuals which should be subjected to this analysis to obtain statistically significant results usually ranges between 30 and 300, with a preferred number of individuals ranging between 50 and 150. The same considerations apply to the number of random control or unaffected individuals used in the study. The results of this first analysis provide haplotype frequencies in case-control populations, the relative risk for an individual carrying a given haplotype of being affected with the given trait under study and the estimated p value for each evaluated haplotype.

Interaction Analysis

The biallelic markers of the present invention may also be used to identify patterns of biallelic markers associated with detectable traits resulting from polygenic interactions. The analysis of genetic interaction between alleles at unlinked loci requires individual genotyping using the techniques described herein. The analysis of allelic interaction among a selected set of biallelic markers with appropriate level of statistical significance can be considered as a haplotype analysis, similar to those described in further details within the present invention. Preferably, genotyping typing is performed using the microsequencing technique.

Methods to test for association between a trait and a biallelic marker allele or a haplotype of biallelic marker alleles are described below.

XI.D. Statistical methods

In general, any method known in the art to test whether a trait and a genotype show a statistically significant correlation is used.

Methods to estimate haplotype frequencies in a population

As described above, when genotypes are scored, it is often not possible to distinguish heterozygotes so that haplotype frequencies cannot be easily inferred. When the gametic phase is not known, haplotype frequencies can be estimated from the multilocus genotypic data. Any method known to person skilled in the art can be used to estimate haplotype frequencies (see Lange K., *Mathematical and Statistical Methods for Genetic Analysis*, Springer, New York, 1997; Weir, B.S., *Genetic data Analysis II: Methods for Discrete population genetic Data*, Sinauer Assoc., Inc., Sunderland, MA, USA, 1996) Preferably, maximum-likelihood haplotype frequencies are computed using an Expectation- Maximization (EM) algorithm (see Dempster et al., *J. R. Stat. Soc.*, 39B:1-38, 1977; Excoffier L. and Slatkin M., *Mol. Biol. Evol.*, 12(5): 921-927, 1995). This procedure is an iterative process aiming at obtaining maximum-likelihood estimates of haplotype frequencies from multi-locus genotype data when the gametic phase is unknown. Haplotype estimations are usually performed by applying the EM algorithm using for example the EM-HAPLO program (Hawley M.E. et al., *Am. J. Phys. Anthropol.*, 18:104, 1994) or the Arlequin program (Schneider et al., *Arlequin: a software for population genetics data analysis*, University of Geneva, 1997). The EM algorithm is a generalized iterative maximum likelihood approach to estimation and is briefly described below.

In the following part of this text, phenotypes will refer to multi-locus genotypes with unknown phase. Genotypes will refer to known-phase multi-locus genotypes.

Suppose a sample of N unrelated individuals typed for K markers. The data observed are the unknown-phase K-locus phenotypes that can be categorized in F different phenotypes. Suppose that we have H underlying possible haplotypes (in case of K biallelic markers, $H=2^K$).

For phenotype j, suppose that c_j genotypes are possible. We thus have the following equation

$$P_j = \sum_{i=1}^{c_j} pr(genotype_i) = \sum_{i=1}^{c_j} pr(h_k, h_l) \quad \text{Equation 1}$$

where P_j is the probability of the phenotype j, h_k and h_l are the two haplotypes constituent the genotype i. Under the Hardy-Weinberg equilibrium, $pr(h_k, h_l)$ becomes :

$$pr(h_k, h_l) = pr(h_k)^2 \text{ if } h_k = h_l, pr(h_k, h_l) = 2 pr(h_k) \cdot pr(h_l) \text{ if } h_k \neq h_l. \quad \text{Equation 2}$$

The successive steps of the E-M algorithm can be described as follows:

Starting with initial values of the of haplotypes frequencies, noted, $p_1^{(0)}, p_2^{(0)}, \dots, p_T^{(0)}$.

these initial values serve to estimate the genotype frequencies (Expectation step) and then estimate another set of haplotype frequencies (Maximization step): $p_1^{(1)}, p_2^{(1)}, \dots, p_T^{(1)}$.

these two steps are iterated until change in the sets of haplotypes frequency are very small.

A stop criterion can be that the maximum difference between haplotype frequencies between two iterations is less than 10^{-7} . These values can be adjusted according to the desired precision of estimations.

In detail, at a given iteration s , the Expectation step consists in calculating the genotypes frequencies by the following equation:

$$\begin{aligned} pr(genotype_i)^{(s)} &= pr(phenotype_j) \cdot pr(genotype_i | phenotype_j)^{(s)} \\ &= \frac{n_j}{N} \cdot \frac{pr(h_k, h_l)^{(s)}}{p_j^{(s)}} \end{aligned} \quad \text{Equation 3}$$

where genotype i occurs in phenotype j , and where h_k and h_l constitute genotype i . Each probability are derived according to equations 1 and 2 above.

Then the Maximization step simply estimates another set of haplotype frequencies given the genotypes frequencies. This approach is also known as gene-counting method (Smith, *Ann. Hum. Genet.*, 21:254-276, 1957).

$$p_i^{(s+1)} = \frac{1}{2} \sum_{j=1}^F \sum_{i=1}^{c_j} \delta_{ji} \cdot pr(genotype_i)^{(s)} \quad \text{Equation 4}$$

where δ_{ji} is an indicator variable which count the number of time haplotype i in genotype j . It takes the values of 0, 1 or 2.

To ensure that the estimation finally obtained are the maximum-likelihood estimations several values of departures are required. The estimations obtained are compared and if they differ the estimations leading to the best likelihood are kept. The term "haplotype determination method" is used to refer to all methods for determinin haplotypes known in the art including expectation-maximization algorithms.

20 Methods to calculate linkage disequilibrium between markers

A number of methods can be used to calculate linkage disequilibrium between any two genetic positions, in practice, linkage disequilibrium is measured by applying a statistical association test to haplotype data taken from a population.

Linkage disequilibrium between any pair of biallelic markers comprising at least one of the biallelic markers of the present invention (M_i, M_j) can be calculated for every allele combination ($M_{i1}, M_{j1}; M_{i1}, M_{j2}; M_{i2}, M_{j1}$ and M_{i2}, M_{j2}), according to the Piazza formula :

$$\Delta M_{i_k, M_{j_l}} = \sqrt{\theta 4 - \sqrt{(\theta 4 + \theta 3)(\theta 4 + \theta 2)}}, \text{ where :}$$

$\theta 4 = - -$ = frequency of genotypes not having allele k at M_i and not having allele l at M_j

$\theta 3 = - +$ = frequency of genotypes not having allele k at M_i and having allele l at M_j

$\theta 2 = + -$ = frequency of genotypes having allele k at M_i and not having allele l at M_j

Linkage disequilibrium (LD) between pairs of biallelic markers (M_i, M_j) can also be calculated for every allele combination ($M_{i1}, M_{j1}; M_{i1}, M_{j2}; M_{i2}, M_{j1}$ and M_{i2}, M_{j2}), according to the maximum-likelihood estimate (MLE) for delta (the composite linkage disequilibrium coefficient), as described by Weir (B.S. Weir, *Genetic Data Analysis*, Sinauer Ass. Eds, 1996). This formula allows linkage
 5 disequilibrium between alleles to be estimated when only genotype, and not haplotype, data are available. This LD composite test makes no assumption for random mating in the sampled population, and thus seems to be more appropriate than other LD tests for genotypic data.

Another means of calculating the linkage disequilibrium between markers is as follows. For a couple of biallelic markers, $M_i (a/b_i)$ and $M_j (a/b_j)$, fitting the Hardy-Weinberg equilibrium, one can
 10 estimate the four possible haplotype frequencies in a given population according to the approach described above.

The estimation of gametic disequilibrium between a_i and a_j is simply:

$$D_{aiaj} = pr(haplotype(a_i, a_j)) - pr(a_i).pr(a_j).$$

Where $pr(a_i)$ is the probability of allele a_i and a_j is the probability of allele a_j . and where
 15 $pr(haplotype(a_i, a_j))$ is estimated as in Equation 3 above.

For a couple of biallelic marker only one measure of disequilibrium is necessary to describe the association between M_i and M_j .

Then a normalized value of the above is calculated as follows:

$$\begin{aligned} D'^{aiaj} &= D_{aiaj} / \max (pr(a_i).pr(a_j), pr(b_i).(b_j)) && \text{with } D_{aiaj} < 0 \\ 20 \quad D'^{aiaj} &= D_{aiaj} / \max (pr(b_i).pr(a_j), pr(a_i).(b_j)) && \text{with } D_{aiaj} > 0 \end{aligned}$$

The skilled person will readily appreciate that other LD calculation methods can be used without undue experimentation.

Linkage disequilibrium among a set of biallelic markers having an adequate heterozygosity rate can be determined by genotyping between 50 and 1000 unrelated individuals, preferably between
 25 75 and 200, more preferably around 100.

Testing for association

Methods for determining the statistical significance of a correlation between a phenotype and a genotype, in this case an allele at a biallelic marker or a haplotype made up of such alleles, is determined by any statistical test known in the art and with any accepted threshold of statistical
 30 significance being required. The application of particular methods and thresholds of significance are well with in the skill of the ordinary practitioner of the art.

Testing for association is performed by determining the frequency of a biallelic marker allele in case and control populations and comparing these frequencies with a statistical test to determine if their is a statistically significant difference in frequency which would indicate a correlation between
 35 the trait and the biallelic marker allele under study. Similarly, a haplotype analysis is performed by estimating the frequencies of all possible haplotypes for a given set of biallelic markers in case and

control populations, and comparing these frequencies with a statistical test to determine if there is a statistically significant correlation between the haplotype and the phenotype (trait) under study. Any statistical tool useful to test for a statistically significant association between a genotype and a phenotype is used. Preferably the statistical test employed is a chi square test with one degree of freedom. A P-value is calculated (the P-value is the probability that a statistic as large or larger than the observed one would occur by chance).

Statistical significance

In preferred embodiments, significance for diagnosis purposes, either as a positive basis for further diagnostic tests or as a preliminary starting point for early preventive therapy, the p value related to a biallelic marker association is preferably about 1×10^{-2} or less, more preferably about 1×10^{-4} or less, for a single biallelic marker analysis and about 1×10^{-3} or less, still more preferably 1×10^{-6} or less and most preferably of about 1×10^{-8} or less, for a haplotype analysis involving several markers. These values are believed to be applicable to any association studies involving single or multiple marker combinations.

The skilled person can use the range of values set forth above as a starting point in order to carry out association studies with biallelic markers of the present invention. In doing so, significant associations between the biallelic markers of the present invention and cancer and prostate cancer can be revealed and used for diagnosis and drug screening purposes.

Using the method described above and evaluating the associations for single marker alleles or for haplotypes permits an estimation of the risk a corresponding carrier has to develop a given trait, and particularly in the context of the present invention, a disease, preferably cancer, more preferably prostate cancer. Significance thresholds of relative risks are to be adapted to the reference sample population used.

In this regard, among all the possible marker combinations or haplotypes which are evaluated to determine the significance of their association with a given trait, for example a form of cancer or prostate cancer, a response to treatment with anti-cancer or anti-prostate cancer agents or side effects related to treatment with anti-cancer or anti-prostate cancer agents, it is believed that those displaying a coefficient of relative risk above 1, preferably about 5 or more, preferably of about 7 or more are indicative of a "significant risk" for the individuals carrying the identified haplotype to develop the given trait. It is difficult to evaluate accurately quantified boundaries for the so-called "significant risk". Indeed, and as it has been demonstrated previously, several traits observed in a given population are multifactorial in that they are not only the result of a single genetic predisposition but also of other factors such as environmental factors or the presence of further, apparently unrelated, haplotype associations. Thus, the evaluation of a significant risk must take these parameters into consideration in order to, in a certain manner, weigh the potential importance of external parameters in the development of a given trait. Without wishing to be bound to any invariable model or theory

based on the above statistical analyses, the inventors believe that a "significant risk" to develop a given trait is evaluated differently depending on the trait under consideration.

It will of course be understood by practitioners skilled in the treatment or diagnosis of cancer and prostate cancer that the present invention does not intend to provide an absolute identification of
5 individuals who could be at risk of developing a particular form of cancer or who will or will not respond or exhibit side effects to treatment with anti-cancer or anti-prostate cancer agents but rather to indicate a certain degree or likelihood of developing a disease or of observing in a given individual a response or a side effect to treatment with a particular agent or set of agents.

However, this information is extremely valuable as it can, in certain circumstances, be used to
10 initiate preventive treatments or to allow an individual carrying a significant haplotype to foresee warning signs such as minor symptoms. In the case of cancer, the knowledge of a potential predisposition, even if this predisposition is not absolute, might contribute in a very significant manner to treatment, or allow for suggestions in changes in diet or the reduction of risky behaviors, e.g. smoking. Similarly, a diagnosed predisposition to a potential side effect could immediately direct
15 the physician toward a treatment, for which such side effects have not been observed during clinical trials.

Phenotypic randomization

In order to confirm the statistical significance of the first stage haplotype analysis described above, it might be suitable to perform further analyses in which genotyping data from case-control
20 individuals are pooled and randomized with respect to the trait phenotype. Each individual genotyping data is randomly allocated to two groups which contain the same number of individuals as the case-control populations used to compile the data obtained in the first stage. A second stage haplotype analysis is preferably run on these artificial groups, preferably for the markers included in the haplotype of the first stage analysis showing the highest relative risk coefficient. This experiment
25 is reiterated between 50 and 200 times, preferably between 75 and 125 times. The repeated iterations allow the determination of the percentage of obtained haplotypes with a significant p-value level below about 1×10^{-3} .

Example 24

Detailed Association Studies

30 The initial association studies between the 8p23 locus and prostate cancer described in Section I.D. were repeated at a higher level of sophistication.

Collection of DNA samples from affected and non-affected individuals

Prostate cancer patients were recruited according to clinical inclusion criteria based on pathological or radical prostatectomy records as described above in Section I. However, the pool of
35 individuals suffering from prostate cancer described in Section I was augmented from the original 185 individuals to a range of between 275 and 491 individuals depending on the marker tested. Similarly,

the control pool of non-diseased individuals described in Section I was augmented from the original 104 individuals to a range of between 130 and 313 individuals depending on the marker tested.

Genotyping Affected and Control Individuals

As for Section I.D., allelic frequencies of the biallelic markers in each population were
5 determined by performing microsequencing reactions on amplified fragments obtained by genomic PCR performed on the DNA samples from each individual as described in Example 5.

Association Studies

Association results were obtained using markers spanning a 650 kb region of the 8p28 locus around PG1 both using single point analysis and haplotyping studies. See Figure 16. As compared
10 with the earlier representation of the initial association results for this region shown in Figure 2, Figure 16 is to scale, since the entire region has now been sequenced. In addition, more markers were generated around the association peak in the area of PG1; each of which has been tested in single point analysis (hence the density of data within this subregion). The haplotyping curve in Figure 16 represents, for each marker considered, the maximum p-value for haplotypes obtained using this
15 marker and any number from all markers harbored by the same BAC and being in Hardy Weinberg Disequilibrium with said marker.

The data presented in Figure 16 shows a strong association between this specific region within 8p23 locus, especially in the area that has been identified as being the PG1 gene, and prostate cancer. The maximum p-value in single point analysis, for the PG1 sub-region, is 3.10^{-3} , while outside
20 of the PG1 subregion, most of the p-values obtained for single point associations are less significant than 1.10^{-1} . The maximum p-value obtained for haplotyping studies is the one obtained for a marker inside PG1's BAC, and equals 3.10^{-6} .

Figure 17 is a graph showing an enlarged view of the single point association results within a 160 kb region comprising the PG1 gene. Markers involved in this enlargement were all located on
25 BAC B0463F01 (see Figure 16), except marker 4-14, which lies in very close proximity, on BAC B0189E08. Figure 17 shows all of the markers which made up the maximum haplotype shown in Figure 16. Some of these markers were later revealed to lie within the promoter, exonic or intronic regions of the PG1 gene. The markers outside the gene were all informative biallelic markers with a least frequent allele present at a frequency of more than 20%, while markers within the gene were a
30 mix of such informative markers and markers whose least frequent allele's frequency is less than 20%. These data confirm and narrow the previous peak of association values seen in Figure 16, to a 40 kb harboring the PG1 gene. Significant associations are obtained for markers starting at the promoter site with marker No. 99-1485, and ending at the 3' UTR site with marker No. 5-66.

Figure 18A is a graph showing an enlarged view of the single point association results of 40
35 kb within the PG1 gene. These data confirm that seven markers within the PG1 gene have one allele associated with prostate cancer, with p-values all similar and more significant than 1.10^{-2} , specifically

markers 99-622 ; 4-77 ; 4-71 ; 4-73 ; 99-598 ; 99-576 ; 4-66. Figure 18B is a table listing the location of markers within PG1 gene, the two possible alleles at each site. For each marker, the disease-associated allele is indicated first ; its frequencies in cases and controls as well as the difference between both are shown ; the odd-ratio and the p-value of each individual marker association are also shown.

The data in Figures 17, 18A, and 18B demonstrate that the markers in the PG1 gene have an association with prostate cancer that is valid, and exhibits similar significance values, regardless whether the considered cases are sporadic or familial cases. Therefore, some PG1 alleles must be general risk factors for any type of prostate cancer, whether familial or sporadic. The fact that several p-values for associated alleles are around 1.10^{-2} suggests that all these markers are in linkage disequilibrium to one another, and can all be used individually to assess PG1 associated prostate cancer susceptibility risk. The prostate cancer associated alleles of the 7 markers discussed above, all exhibit an odd-ratio of about 1.5, which means for each of them that an individual carrying such allele has 1.5 more chances to be susceptible to prostate cancer than not.

In order to confirm the significance of the association results found for markers on the BAC harboring PG1, we a novel statistical method was performed as described in provisional patent application serial no. 60/107,986, filed November 10, 1998.

Haplotype analysis

The results of a haplotype analysis study using 4 markers (marker Nos. 4-14, 99-217, 4-66 and 99-221) within the 160 kb region shown in Figure 17 are shown in Figure 19A. These 4 markers have each been shown to be strongly associated with prostate cancer, i.e. with p-values more significant than 1.10^{-3} on approximately 150 cases and 130 controls. All haplotypes using 2, 3, or 4 markers among the 4 above cited were analyzed using 491 case patients and 317 control individuals. Figure 19A shows the most significant haplotypes obtained, as well as the individual odd-ratios for each. Haplotype 11 is the most significant (p-value of ca. 3.10^{-6}), and is related to haplotype 5, shown in Figure 4 in that three of the four marker alleles (4-14 C, 99-217 T and 99-221 A) are common to both haplotypes, and both cover a similar region. Differences in p-values are explained both by the addition of markers and of more case or control individuals. Haplotype 11 has an highly informative odd-ratio (of above 3) ; it is present in 3% of the controls and almost 10% of the cases.

Figure 19B is a table showing the segmented haplotyping results according to the age of the subjects, and whether the prostate cancer cases were sporadic or familial, using the same markers 4 markers and the same individuals as were used to generate the results in Figure 19A. Figure 19B shows equivalent results for all segments of the population analyzed, demonstrating that the PG1 associated alleles are general risk factors for prostate cancer, regardless of the age of onset of the disease.

The haplotyping results and odd ratios for all of the combinations of the 7 markers (99-622; 4-77; 4-71; 4-7 ; 99-598; 99-576; and 4-66) within PG1 gene that were shown in Figure 18 to have p-values more significant than 1×10^{-2} were computed. A portion of these data are shown in Figure 20. All of the 2-, 3-, 4-, 5-, 6- and 7-marker haplotypes were tested. Figure 20 identifies for each x-marker haplotype category, the most significant haplotype. Among all these, the most significant haplotype is the two-marker haplotype 1, which shows a p-value of approximately 6.10^{-5} ; with an odd ratio of 2. The frequency of haplotype 1 among the control individuals is 15%, while it is 26% among the case patients. It is worth noting that these frequencies are very similar for all haplotypes presented on Figure 20. It will thus be sufficient to test this two marker haplotype for prognosis/diagnosis on risk patients, as opposed to having a more complex test of a haplotype comprising 3 or more makers.

Finally, Figure 21 is a graph showing the distribution of statistical significance, as measured by Chi-square values, for each series of possible x-marker haplotypes, (x =2, 3 or 4) using all of the 19 markers found in PG1 gene. These data confirm that testing 2-marker haplotypes within PG1 is sufficient because the testing 3- or 4-marker haplotypes does not increase the statistical relevance of the analysis.

Example 25

Attributable Risk

Attributable risk describes the proportion of individuals in a population exhibiting a phenotype due to exposure to a particular factor. For further discussion of attributable risk values, see Holland, Bart K., *Probability without Equations – Concepts for Clinicians*; The Johns Hopkins University Press, pp. 88-90. In the present case the phenotype examined was prostate cancer, and the exposure was either one single allele of an individual PG1-related marker, or a haplotype thereof in an individual's genome.

The formula used for calculating attributable risk values in the present study was the following:

$$AR = P_E(RR-1) / [P_E(RR-1)+1], \text{ where:}$$

AR was the attributable risk of allele or haplotype ;

P_E was the frequency of exposure to allele or haplotype within the population at large, in the present study a random male Caucasian population ; and

RR was the relative risk, in the present study relative risk is approximated with the odd-ratio, because of the relatively low incidence of prostate cancer in populations at large (values for the odd ratios are found in Figures 18B and 20).

In this case, P_E was estimated using a dominant transmission model for prostate cancer:

$$P_E = (N_{AA} + N_{AB}) / N, \text{ where:}$$

N_{AA} was the number of homozygous individuals harboring the disease associated allele or haplotype within a given random population, and N_{AB} was the number of heterozygous individuals in said random population. N_{AA} and N_{AB} were calculated using the allele frequencies in the random

population as indicated in Figures 18B and 20, and N was the number of individuals in total random population.

We calculated the attributable risks of disease-associated alleles for markers within PG1 gene and presented these results in Figure 18B. In Figure 20, the attributable risk for the two-marker
5 haplotypes present in the figure as shown as well. These data demonstrate that disease-associated alleles of PG1 are present in approximately 20% of prostate cancer patients in the Caucasian population at large, and therefor represent prognostic tools of significant value.

SEQUENCE LISTING FREE TEXT

10 The following free text appears in the accompanying Sequence Listing:

identification method Proscan

potential start codon

exon1

Tyr phos

15 upstream amplification primer

polymorphic fragment

polymorphic base

downstream amplification primer

complement

20 upstream amplification primer 99-217-PU, extracted from SEQ ID1 34216 34234

Klein, Kanehisa and DeLisi identification method, potential helix

Eisenberg, Schwarz, Komarony, Wall identification method, potential helix

Prosite match

potential Tyrosine kinase site, Prosite match

25 potential caseine kinase II site, Prosite match

potential Leucine zipper site, Prosite match

potential site, Prosite match

potential protein kinase C, Prosite match

potential cAMP and cGMP dependant protein kinase site, Prosite match

30 primer oligonucleotide

box2 from SEQID4, present in AF003136, P33333, P26647, U89336, U56417,

AB005623

box2 from Z72511

box3 from SEQID4, present in AF003136

35 potential microsequencing oligo

complement potential microsequencing oligo

- polymorphic fragment 4-77, extracted from SEQ ID1 12057 12103
polymorphic fragment 99-123, variant version of SEQ ID21
base A ; G in SEQ ID22
downstream amplification primer 99-217-RP, extracted from SEQ ID1 34625 34645
- 5 complement
polymorphic base C in PG1 (13680) SEQ ID1
stop codon
potential
amplification oligonucleotide
- 10 sequencing oligonucleotide
Box II
Box III
Box I
upstream amplification primer for SEQ 188, SEQ 265, SEQ 189, SEQ 266
- 15 downstream amplification primer for SEQ 185, SEQ 262, SEQ 186, SEQ 263, SEQ 187, SEQ
264 microsequencing oligo for 4-20-149.mis1

Although this invention has been described in terms of certain preferred embodiments, other
20 embodiments which will be apparent to those of ordinary skill in the art in view of the disclosure
herein are also within the scope of this invention. Accordingly, the scope of the invention is intended
to be defined only by reference to the appended claims.

Table 1

marker	polymorphism	most frequent	less frequent	cases		controls	
				p *	q **	p *	q **
99-123	C/T	C	T	0,65	0,35	0,7	0,3
4-26	A/G	A	G	0,61	0,39	0,55	0,45
4-14	C/T	C	T	0,65	0,35	0,59	0,41
4-77	C/G	C	G	0,67	0,33	0,76	0,24
99-217	C/T	C	T	0,69	0,31	0,77	0,23
4-67	C/T	C	T	0,74	0,26	0,84	0,16
99-213	A/G	A	G	0,55	0,45	0,62	0,38
99-221	C/A	C	A	0,43	0,57	0,43	0,57
99-135	A/G	A	G	0,75	0,25	0,7	0,3

*: frequency of most frequent base within each sub-population

**: frequency of least frequent base within each sub-population ($p+q=1$)

standard deviations -0,023 to 0,031 for controls

standard deviations -0,018 to 0,021 for cases

WHAT IS CLAIMED IS:

1. A recombinant, purified or isolated polynucleotide comprising a mammalian PG1 gene, cDNA, complement thereof, or fragment thereof having at least 10 nucleotides in length.
5
2. The polynucleotide according to claim 1, wherein said mammalian PG1 gene or cDNA is human or mouse.
3. The polynucleotide according to claim 2, wherein the polynucleotide is selected from
10 SEQ ID NOs: 3, 69, 112-124, 179, and 182-184.
4. A polynucleotide selected from SEQ ID NOs: 185-578.
5. A purified or isolated polypeptide comprising a mammalian PG1 protein, or fragment
15 thereof having at least 8 amino acids in length.
6. The polypeptide according to claim 5, wherein said mammalian PG1 protein is human or mouse.
- 20 7. The polypeptide according to claim 6, wherein said polypeptide is selected from SEQ ID NOs: 4, 5, 70, 74, and 125-136.
8. The polypeptide according to claim 5, wherein said polypeptide consists of said mammalian PG1 protein, or fragment thereof having at least 8 amino acids in length.
25
9. A polynucleotide comprising a nucleic acid sequence encoding a polypeptide according to claim 8.
10. An antibody composition capable of selectively binding to an epitope-containing
30 fragment of a polypeptide according to claim 8, wherein said antibody is either polyclonal or monoclonal.
11. A vector comprising a polynucleotide according to any one of claims 1, 4, and 9.
- 35 12. A host cell comprising a polynucleotide according to claim 11.

- 13 A nonhuman host animal or mammal comprising a vector according to claim 11.
14. A mammalian host cell comprising a PG1 gene disrupted by homologous recombination with a knock out vector.
- 5 15. A nonhuman host mammal comprising a PG1 gene disrupted by homologous recombination with a knock out vector.
16. A polynucleotide according to any one of claims 1, 4, and 9, further comprising a
10 label.
17. A polynucleotide according to any one of claims 1, 4, and 9, attached to a solid support.
- 15 18. A random or addressable array of polynucleotides comprising at least one polynucleotide according to any one of claims 1, 4, and 9.
19. A method of determining whether an individual is at risk of developing cancer or prostate cancer, or whether said individual suffers from cancer or prostate cancer as a result of a
20 mutation in the PG1 gene comprising:
obtaining a nucleic acid sample from said individual; and
determining whether the nucleotides present at one or more PG1-related biallelic marker are indicative of a risk of developing cancer or prostate cancer or indicative of cancer or prostate cancer resulting from a mutation in the PG1 gene.
- 25 20. A method of determining whether an individual is at risk of developing cancer or prostate cancer or whether said individual suffers from cancer or prostate cancer as a result of a mutation in the PG1 gene comprising:
obtaining a nucleic acid sample from said individual; and
30 determining whether the nucleotides present at one or more PG1-related biallelic marker are indicative of a risk of developing cancer or prostate cancer or indicative of cancer or prostate cancer resulting from a mutation in the PG1 gene.
21. A method according to either one of claims 19 and 20, wherein said PG1-related
35 biallelic is a PG1-related biallelic markers positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected from the group consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-

77/151, 4-71/233, 4-72/127, 4-73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-600/492, 99-598/130, 99-217/277, 99-576/421, 4-61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576, and 4-66.

5 22. A method of obtaining an allele of the PG1 gene which is associated with a detectable phenotype comprising:

obtaining a nucleic acid sample from an individual expressing said detectable phenotype;

contacting said nucleic acid sample with an agent capable of specifically detecting a nucleic acid encoding the PG1 protein; and

10 isolating said nucleic acid encoding the PG1 protein.

23. A method of obtaining an allele of the PG1 gene which is associated with a detectable phenotype comprising:

obtaining a nucleic acid sample from an individual expressing said detectable phenotype;

15 contacting said nucleic acid sample with an agent capable of specifically detecting a sequence within the 8p23 region of the human genome;

identifying a nucleic acid encoding the PG1 protein in said nucleic acid sample; and

isolating said nucleic acid encoding the PG1 protein.

20 24. A method of categorizing the risk of prostate cancer in an individual comprising the step of assaying a sample taken from the individual to determine whether the individual carries an allelic variant of PG1 associated with an increased risk of prostate cancer.

25 25. The method of Claim 24 wherein said sample is a nucleic acid sample.

26. The method of Claim 24 wherein said sample is a protein sample.

27. The method of Claim 26, further comprising determining whether the PG1 protein in said sample binds an antibody that binds specifically to a PG1 isoform associated with prostate
30 cancer.

28. A method of genotyping comprising determining the identity of a nucleotide at a PG1-related biallelic marker in a biological sample.

29. A method of estimating the frequency of an allele in a population comprising determining the proportional representation of a nucleotide at a PG1-related biallelic marker in a pooled biological sample derived from said population.

5 30. A method of detecting an association between a genotype and a phenotype, comprising the steps of:

- a) genotyping at least one PG1-related biallelic marker in a trait positive population;
 - b) genotyping said PG1-related biallelic marker in a control population; and
 - c) determining whether a statistically significant association exists between said genotype and
- 10 said phenotype.

31. A method of estimating the frequency of a haplotype for a set of biallelic markers in a population, comprising:

- a) genotyping at least one PG1-related biallelic marker;
- 15 b) genotyping a second biallelic marker by determining the identity of the nucleotides at said second biallelic marker for both copies of said second biallelic marker present in the genome of each individual in said population; and
- c) applying an haplotype determination method to the identities of the nucleotides determined in steps a) and b) to obtain an estimate of said frequency.

20

32. A method of detecting an association between a haplotype and a phenotype, comprising the steps of:

- a) estimating the frequency of at least one haplotype in a trait positive population according to the method of claim 31;
- 25 b) estimating the frequency of said haplotype in a control population according to the method of claim 31; and
- c) determining whether a statistically significant association exists between said haplotype and said phenotype.

30 33. A method according to claim 31, wherein said PG1-related biallelic marker and said second biallelic marker are 4-77/151 and 4-66/145,

34. A method according to claim 32, wherein said haplotype exhibits a p-value of $< 1 \times 10^{-3}$ in an association with a trait positive population with cancer, or prostate cancer.

35

35. A method according to any one of claims 29 to 31, wherein said PG1-related biallelic is a PG1-related biallelic markers positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected from the group consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-77/151, 4-71/233, 4-72/127, 4-73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-600/492, 99-598/130, 5 99-217/277, 99-576/421, 4-61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576, and 4-66.

36. A method according to either one of claims 30 and 32, wherein said control population is a trait negative population or a random population.

10

37. A method according to any one of claims 22, 23, 30, and 32, wherein said phenotype is a disease, cancer or prostate cancer; a response to an anti-cancer agent or an anti-prostate cancer agent; or a side effect to an anti-cancer or anti-prostate cancer agent.

15

38. A polynucleotide for use in a hybridization assay for determining the identity of the nucleotide at an PG1-related biallelic marker; for use in a sequencing assay for determining the identity of the nucleotide at an PG1-related biallelic marker; for use in a allele-specific amplification assay for determining the identity of the nucleotide at an PG1-related biallelic marker; or for use in amplifying a segment of nucleotides comprising an PG1-related biallelic marker.

20

39. The polynucleotide according to claim 38, wherein said PG1-related biallelic is a PG1-related biallelic markers positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected from the group consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-77/151, 4-71/233, 4-72/127, 4-73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-600/492, 99-598/130, 99-217/277, 99-576/421, 4-61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected 25 from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576, and 4-66.

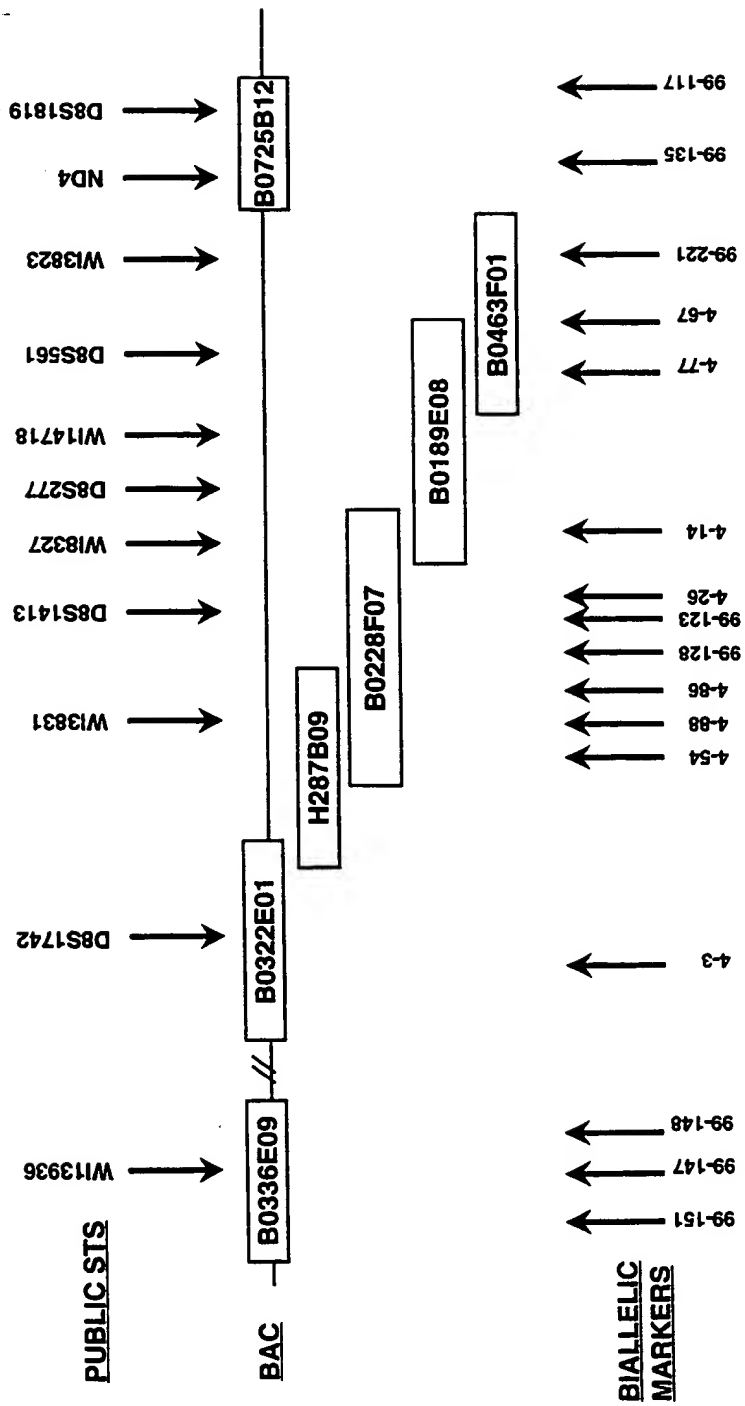


Figure 1

Figure 2

ASSOCIATION STUDIES (FIRST SCREENING)

Population	AFFECTED	NON AFFECTED
Sample size	CASES = 112	CONTROLS=76
Population	35 sporadic cases	> 65 years
Characteristics	+ 77 familial cases	PSA<4

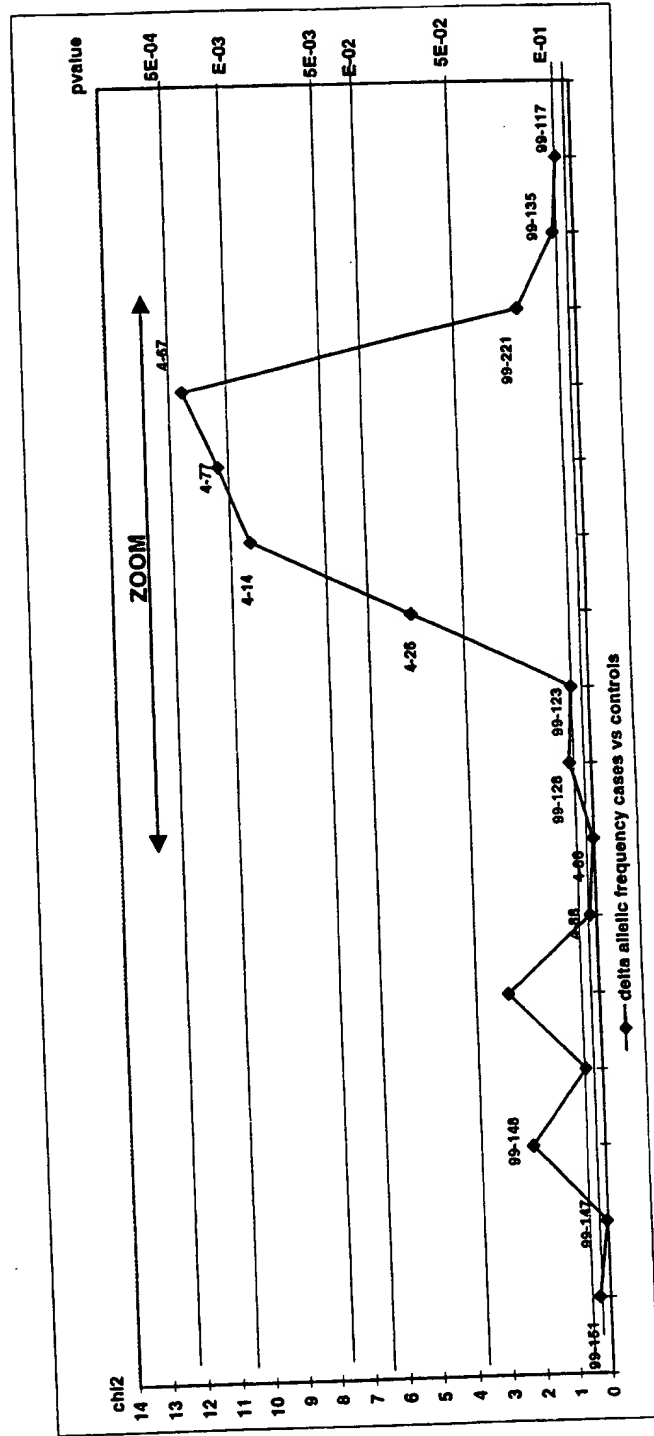


Figure 3

ASSOCIATION STUDIES (ZOOM)

characteristics of populations	AFFECTED CASES (185)	UNAFFECTED CONTROLS (104)
	47 sporadic cases + 138 familial cases	> 65 years PSA<4

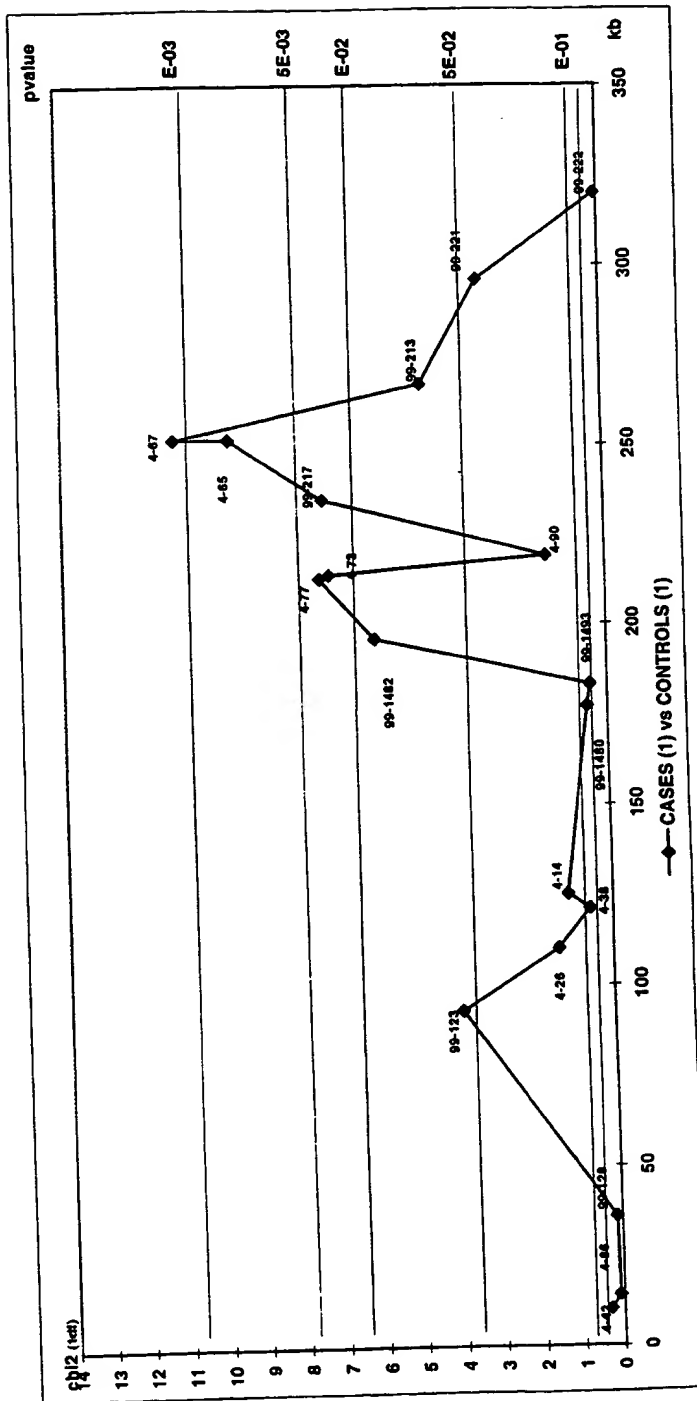


Figure 5

HAPLOTYPE SIMULATIONS (100 ITERATIONS)

markers	4-14	4-77	99-217	4-67	99-213	99-221	haplotype frequencies		relativ risk	pvalue
							cases	controls		
haplotype	C	G	T	T	G	A	0,117	0,013	10,06	9,00E-07

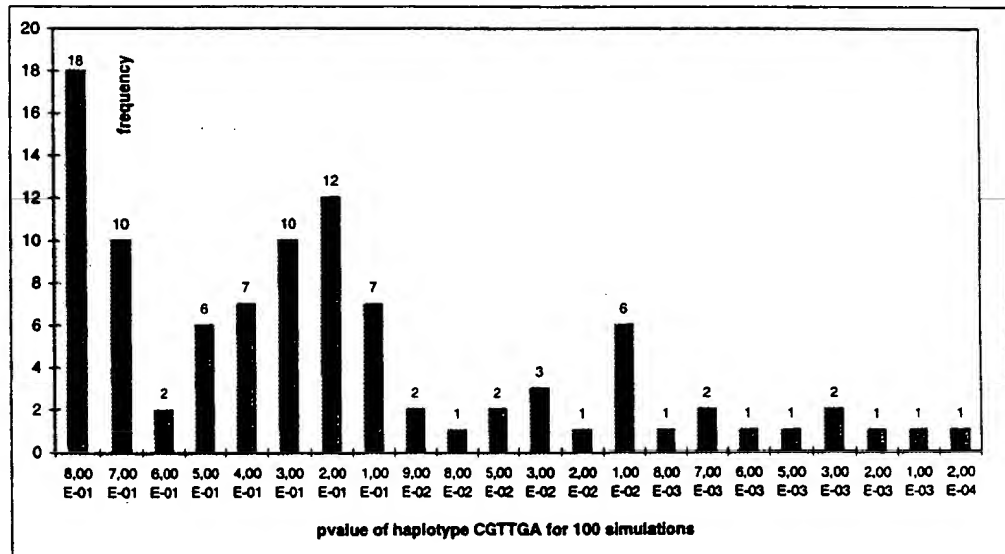
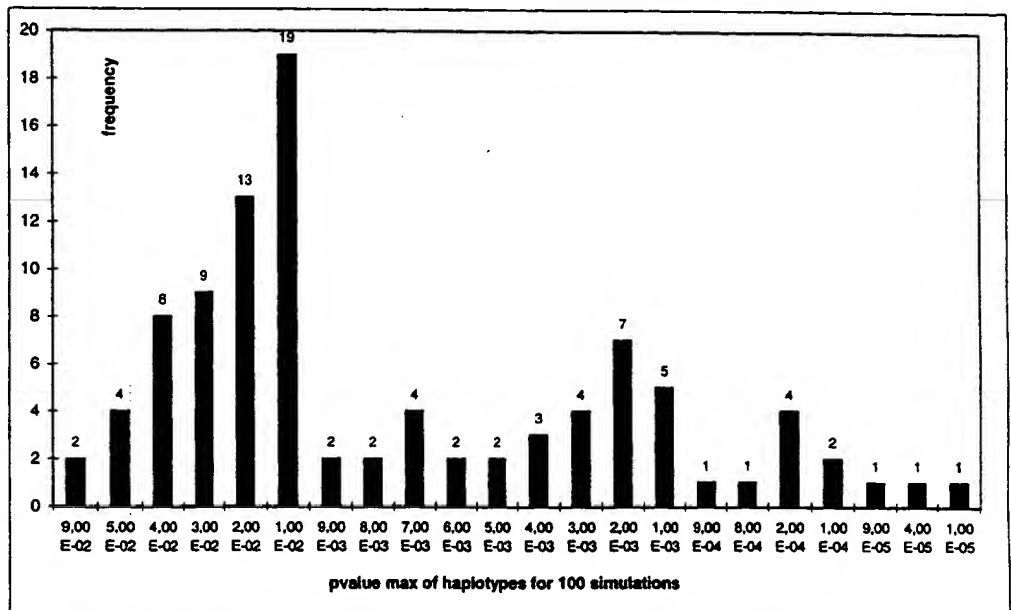


Figure 6A

BAC	Marker	SEQ ID N°	SEQ ID N° (mut)	PU sequence	SEQ ID N°	RP sequence	SEQ ID N°	Polymorphism position*	Base	Microseq. oligos positions*
228	99-123	21	30	AAAGCCAGGACTAGAGG	39	TATTCAGAAAGGAGTGGG	48	24	C/T	1-23 25-47 (complementary)
189	4-26	22	31	TACAGCCCTGTAAAGACAC	40	TGAGGACTGCTAGGAAAG	49	24	A/G	1-23 25-47 (complementary)
228/189	4-14	23	32	TCTAACCTCTCATCCAAC	41	GACTGTATCCTTTGATGCAC	50	24	C/T	1-23 25-47 (complementary)
189/463	4-77	24	33	TGTTGATTTACAGCGCGC	42	GAAAGGACTCATTCATAG	51	24	G/C	1-23 25-47 (complementary)
463	99-217	25	34	GGTGGAAATTTACTATATG	43	GTTTATTTTGTGTGAGCTTGG	52	24	C/T	1-23 25-47 (complementary)
189/463	4-67	26	35	AAGTCCACCTTCTCAAGC	44	TGAAAGATTTATCTCTGG	53	24	C/T	1-23 25-47 (complementary)
463	99-213	27	36	ATACTGCGCAGCGTGTGCTTC	45	TTATGCCCCACATGCTGAG	54	24	C/T	1-23 25-47 (complementary)
463	99-221	28	37	CCCTTTTCTTCAGTGTTC	46	TCATTGCTGTGGCTAGGTC	55	24	A/C	1-23 25-47 (complementary)
725	99-135	29	38	TGGAAGTTGTATTGCCC	47	AAACACCTCCCATTTGTC	56	24	A/G	1-23 25-47 (complementary)

Figure 6B

BAC	Marker	SEQ ID N°	SEQ ID N° (mut)	PU sequence	SEQ ID N°	RP sequence	SEQ ID N°	Polymorphism position*	Base	Microseq. oligos positions*
189/463	99-1482	57	60	ATCAATCAGTGAAGTCTGAG	63	ACAAATCTATATAAGGCTGG	66	24	A/C	1-23 25-47 (complementary)
463	4-73	58	61	ATCGCTGGAACATCTGCG	64	CTCTGGTTAAACAGCAGTG	67	24	G/C	1-23 25-47 (complementary)
463	4-65	59	62	GATTTAAGTACGCTATTAG	65	TGGCTCTGCATTCTTCC	68	24	C/T	1-23 25-47 (complementary)

*: positions are given relative to the sequence of the corresponding marker (i.e. SEQ ID N° 21-38 and 57-62)

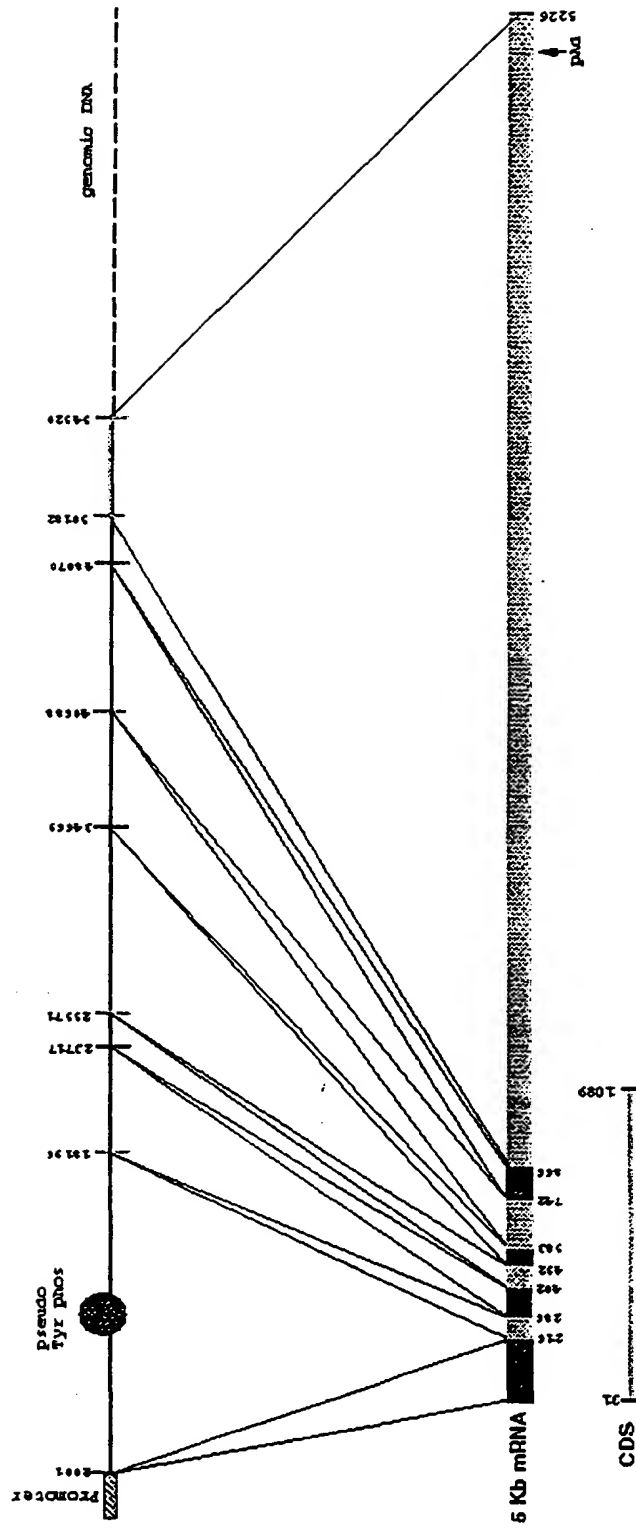
Figure 7

EXON Phase	START	END	5' SPsite	PHASE	3' SPsite
Ex1 +0	2001	2216			GTGAGC
Ex2 +1 GTTTGTA	18196	18265	TAG	+0	
Ex3 +0	23717	23832	CAG	+2	GTAAGT
Ex4 +0 GTAAGA	25571	25660	CAG	+0	
Ex5 +2 GTAAGT	34669	34759	CAG	+0	
Ex6 +1	40688	40846	TAG	+1	GTAAGT
Ex7 +2	48070	48193	TAG	+2	GTGAGT
Ex8	50182	54523	TAG	+1	
ATG codon	2031	2033			
STOP codon	50405	50407			
POLY Ad site	54445	54450			

8/24

PG1 genome structure and cDNAs

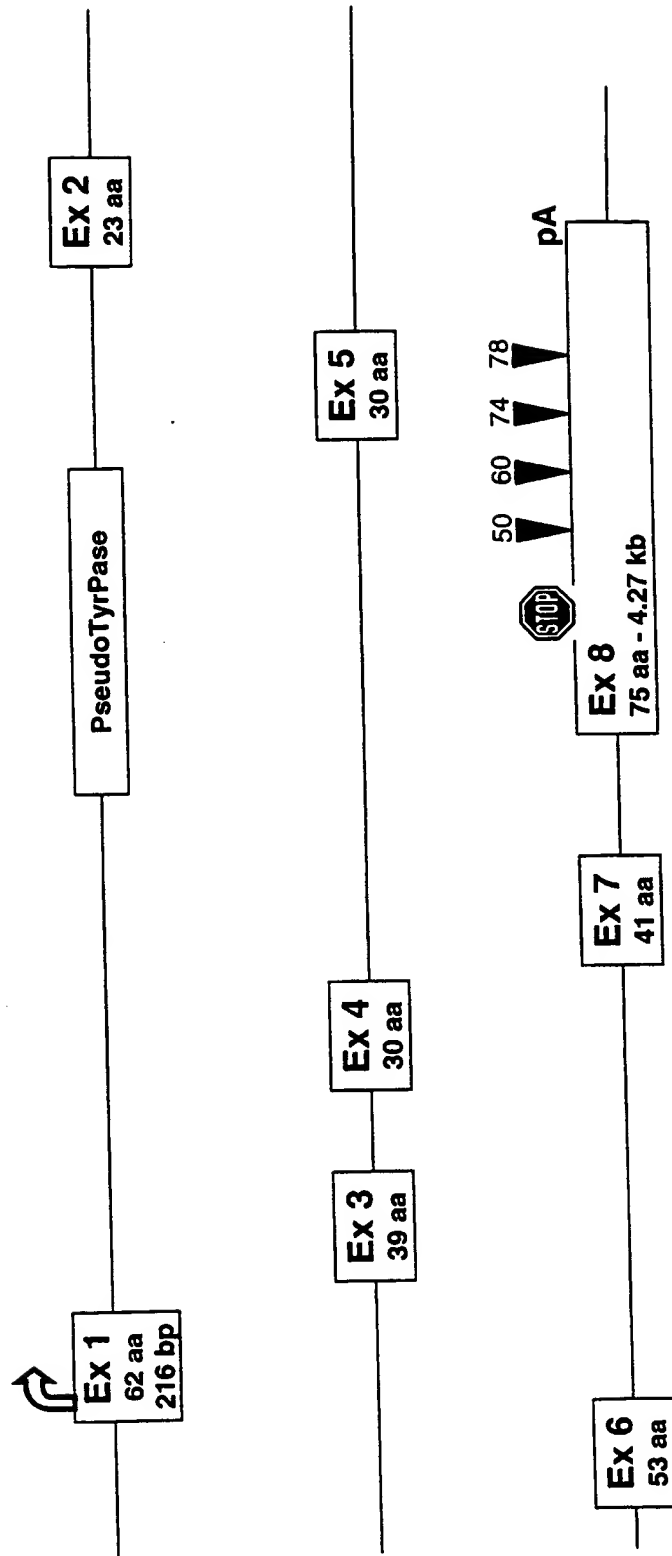
Figure 8A



9/24

Prostate Gene 1

Figure 8B



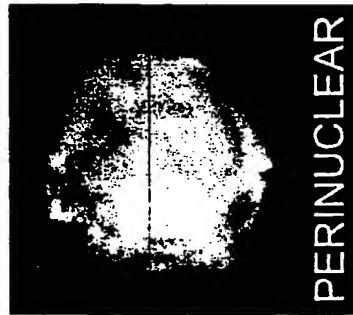
		box1	box2	box3
PG1	Hs	NHQ 81-83	FPEGTR 160-165	LDAIYDVTV 211-219
AF003136 (Genbank)	Ce	NHQ 630-632	FPEGTR 712-717	LDAIYDVTV 762-770
Z72511 (Genbank)	Ce	48 NHR 50	FPEGTD 129-134	VEITYDITI 204-212
P38226 (Swissprot)	Sc	111 NHQ 113	FPEGTN 223-228	IESLYDITI 271-279
P33333 (Swissprot)	Sc	81 NHQ 83	FPEGTR 154-159	-
Z49770 (Genbank)	Sc	116 NHQ 118	FPEGTN 215-220	LDAIYDVTI 265-273
P26647 (Swissprot)	Ec	72 NHQ 74	FPEGTR 145-150	-
Z49860 (Genbank)	Bn	-	FVEGTR 90-95	VPAIYDMTV 138-146
U89336 (Genbank)	Hs	95 NHQ 97	FPEGTR 168-173	-
U56417 (Genbank)	Hs	103 NHQ 105	FPEGTR 176-181	-
AB005623 (Genbank)	Mm	100 NHQ 102	FPEGTR 173-178	-
Z29518 (Genbank)	Zm	91 NHR 93	FVEGTR 170-175	VPAIYDTTV 218-226

Hs = Homo sapiens, Ce = Caenorabibtis elegans, Ec = Escherichia coli; Sc = Saccharomyces cerevisiae, Bn = Brassica napus, Zm = Zea maize, Mm = Mus musculus

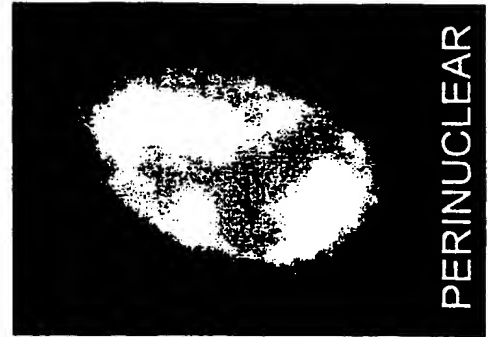
- = pattern absent from protein sequence

Note : Functional acyl glycerol transferases all contain boxes 1 and 2 and not box 3.
Proteins most related to PG1 contain the 3 boxes with a high degree of conservation.

Figure 9



PC3
PG1



PNT2
PG1

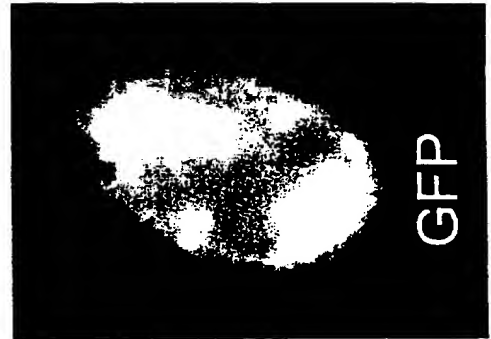


Figure 10

Figure 11

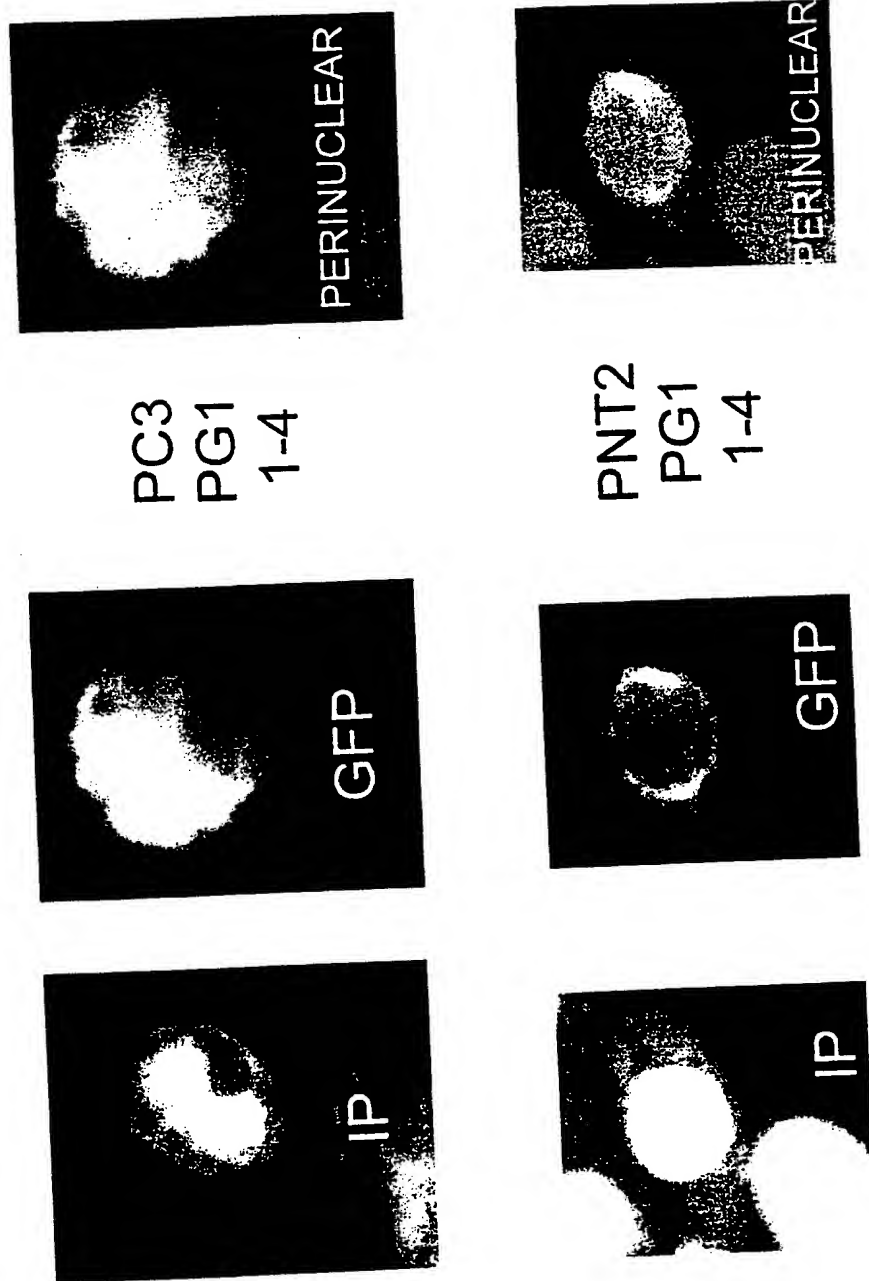
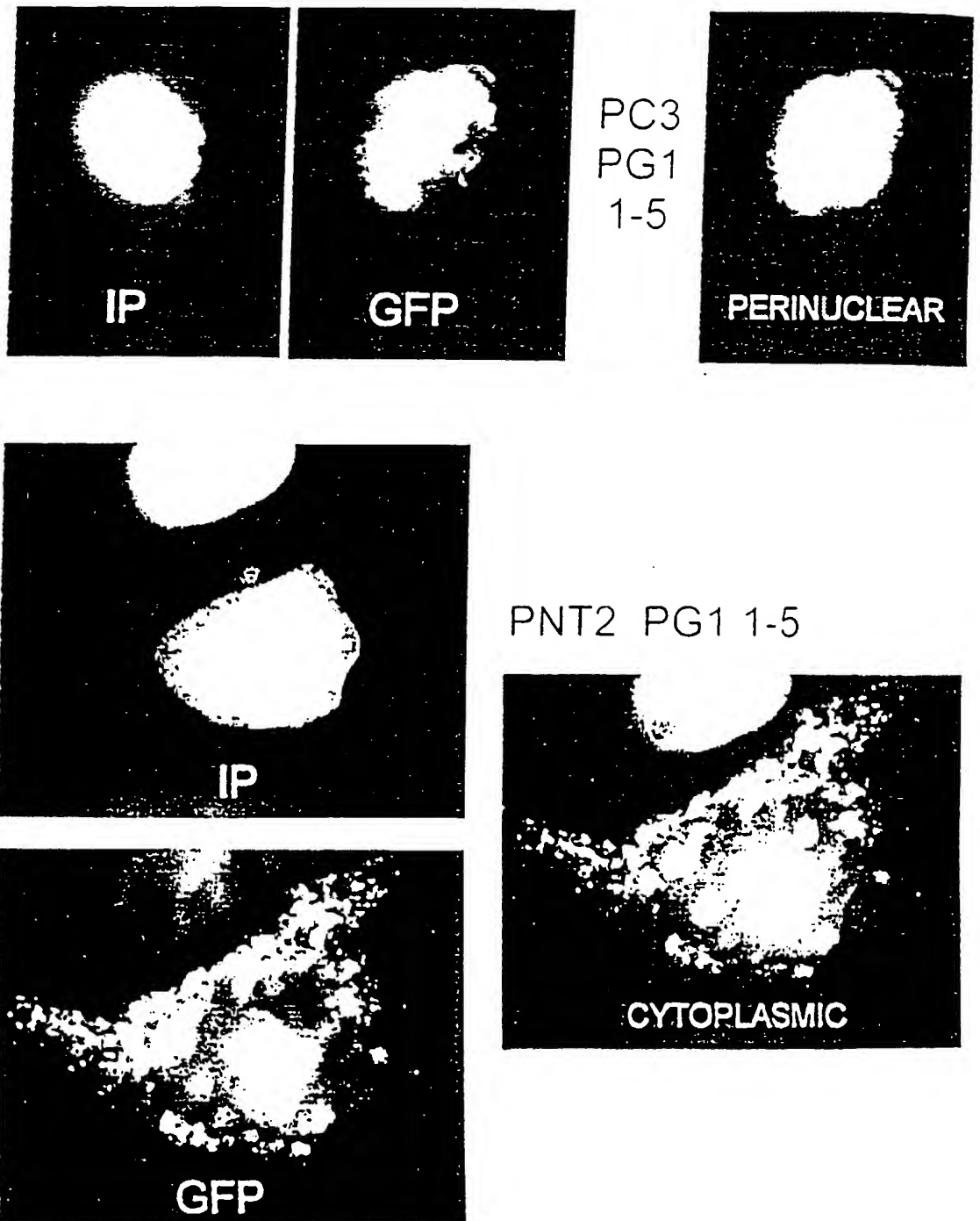
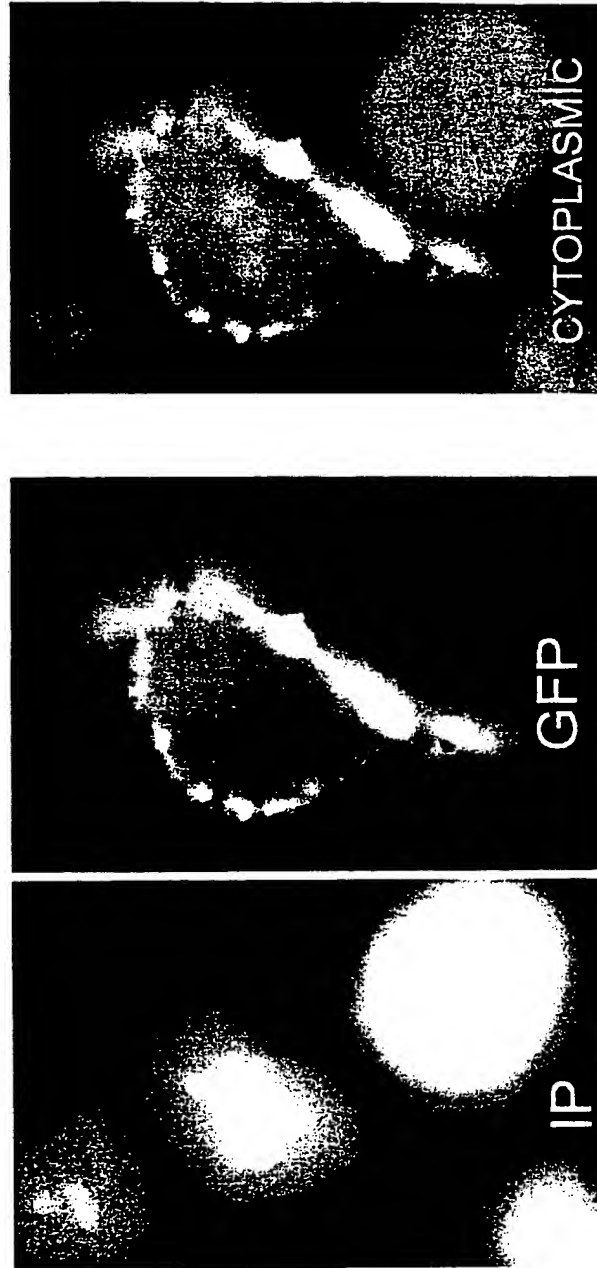


Figure 12





PNT2 PG1 mut229

Figure 13

Figure 14

Alternative splicing

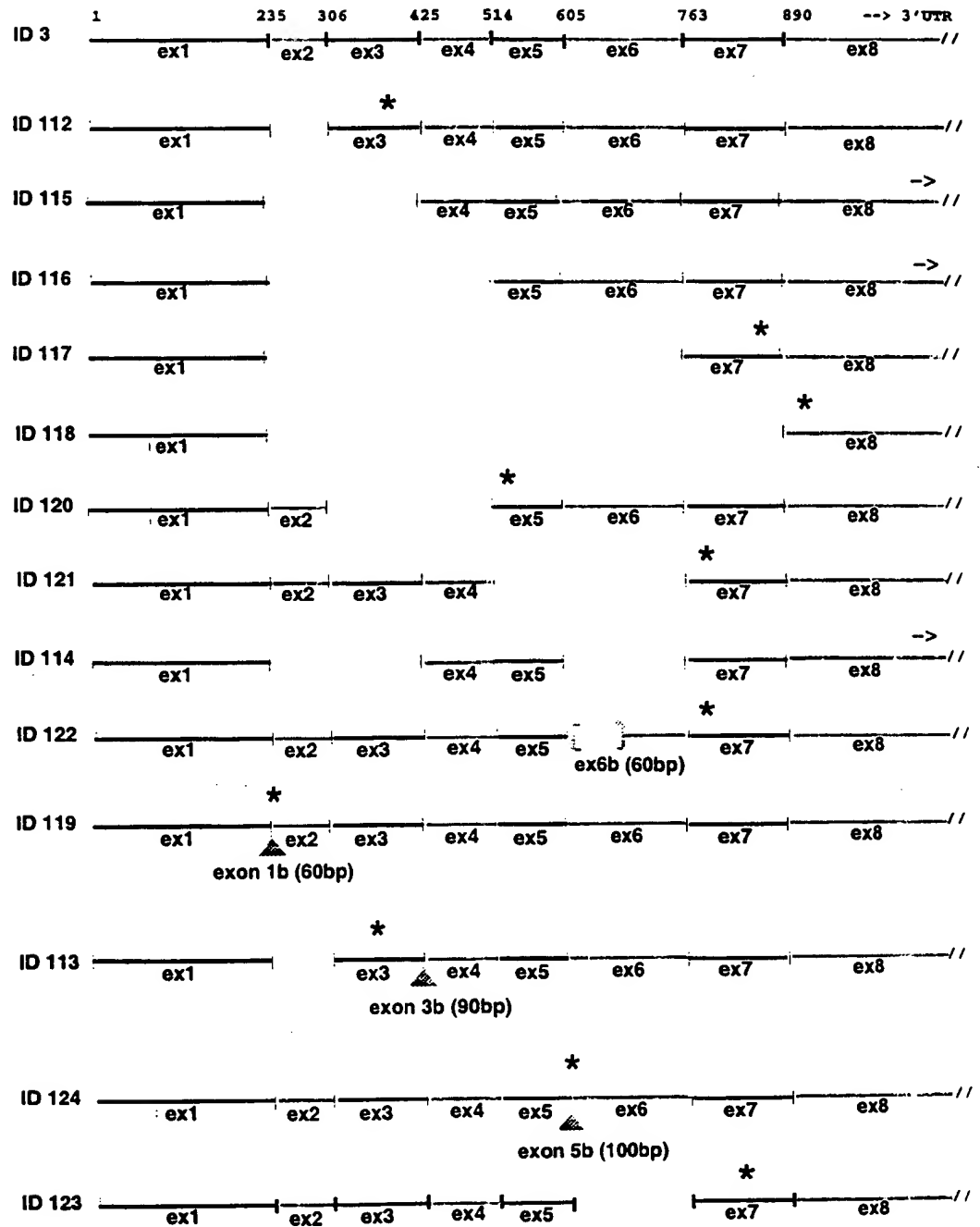


Figure 15

16/24

Combinations of exons of PG1 gene discovered by PCR with primers specific for exon borders

Primer	Clones	prostate	PNT1A	PNT1B	PNT2	LaCaPFCG	LNCaPMB	CoHPV	Du145	PC3	ECP5	ECP6	ECP7	ECP8	ECP9	ECP10	ECP11	ECP12	ECP13	ECP14	ECP15	ECP16	ECP17	ECP18	ECP19	ECP20	ECP21	ECP22	ECP23	ECP24
PGlexon13	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon14	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon15	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon17	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon18	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon24	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon25	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon26	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon27	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon28	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon35	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon36	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon37	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon38	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon46	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon47	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon48	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon57	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon58	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon68	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon11b	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon1b2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon1b3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon1b4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon1b5	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon1b6	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon1b7	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon1b8	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon3b4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon3b5	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon3b6	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon3b7	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon3b8	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon5b6	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon5b7	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon5b8	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon56b	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon46b	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon36b	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon26b	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PGlexon16b	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

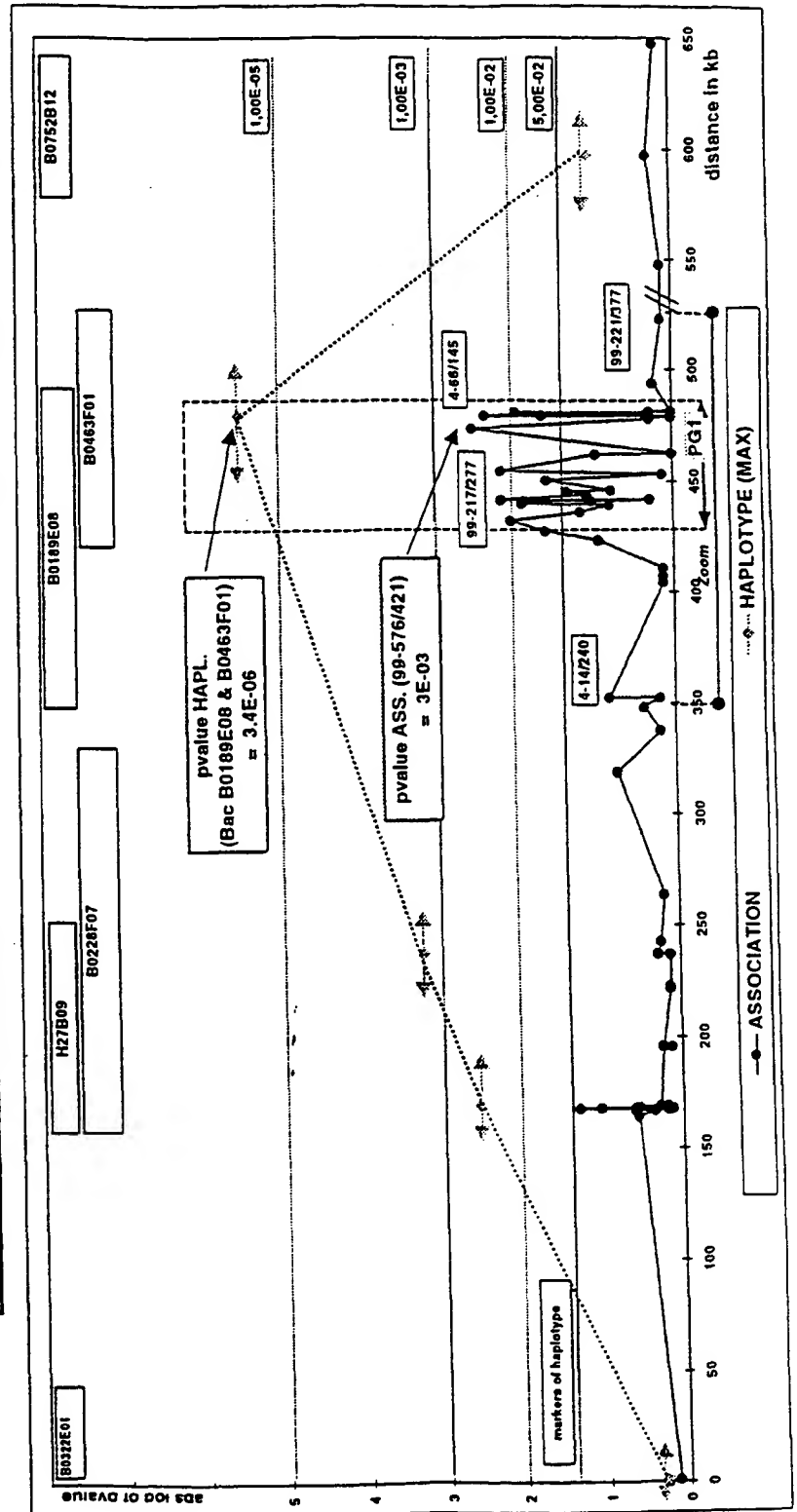
[+] alternative splicing form with combination of exons 13478 instead of 1345678

17/24

ASSOCIATION STUDIES HAPLOTYPE FREQUENCY ANALYSIS (chrom 8p23)

Figure 16

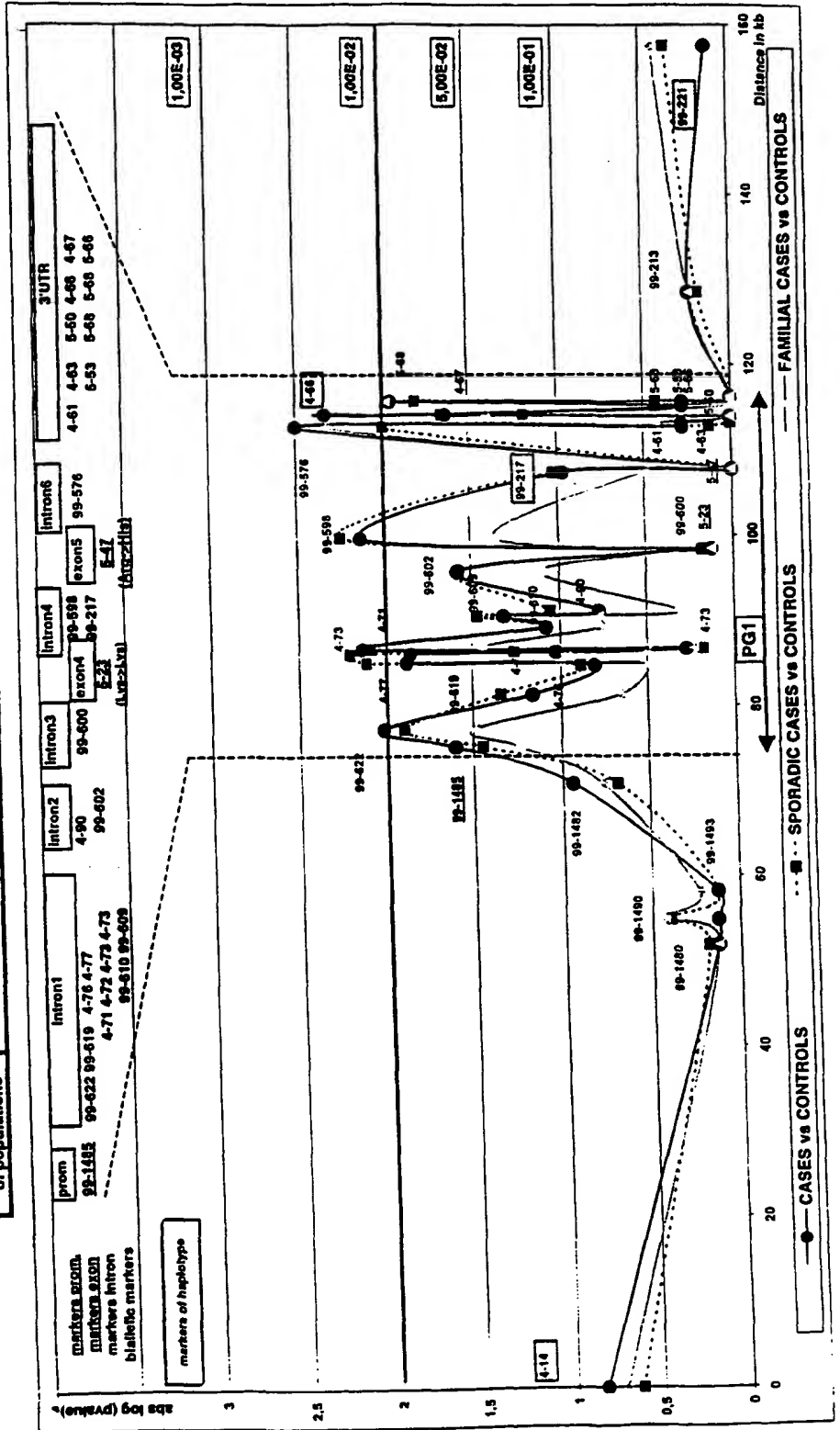
POPULATIONS	AFFECTED N= [275;491]	UNAFFECTED N=[130;313]
	ALL CASES	CONTROLS
characteristics of populations	[275, 491] all cases	<28 unaffected (65 years or older) & <289 controls (65 years or older with PSA<4)

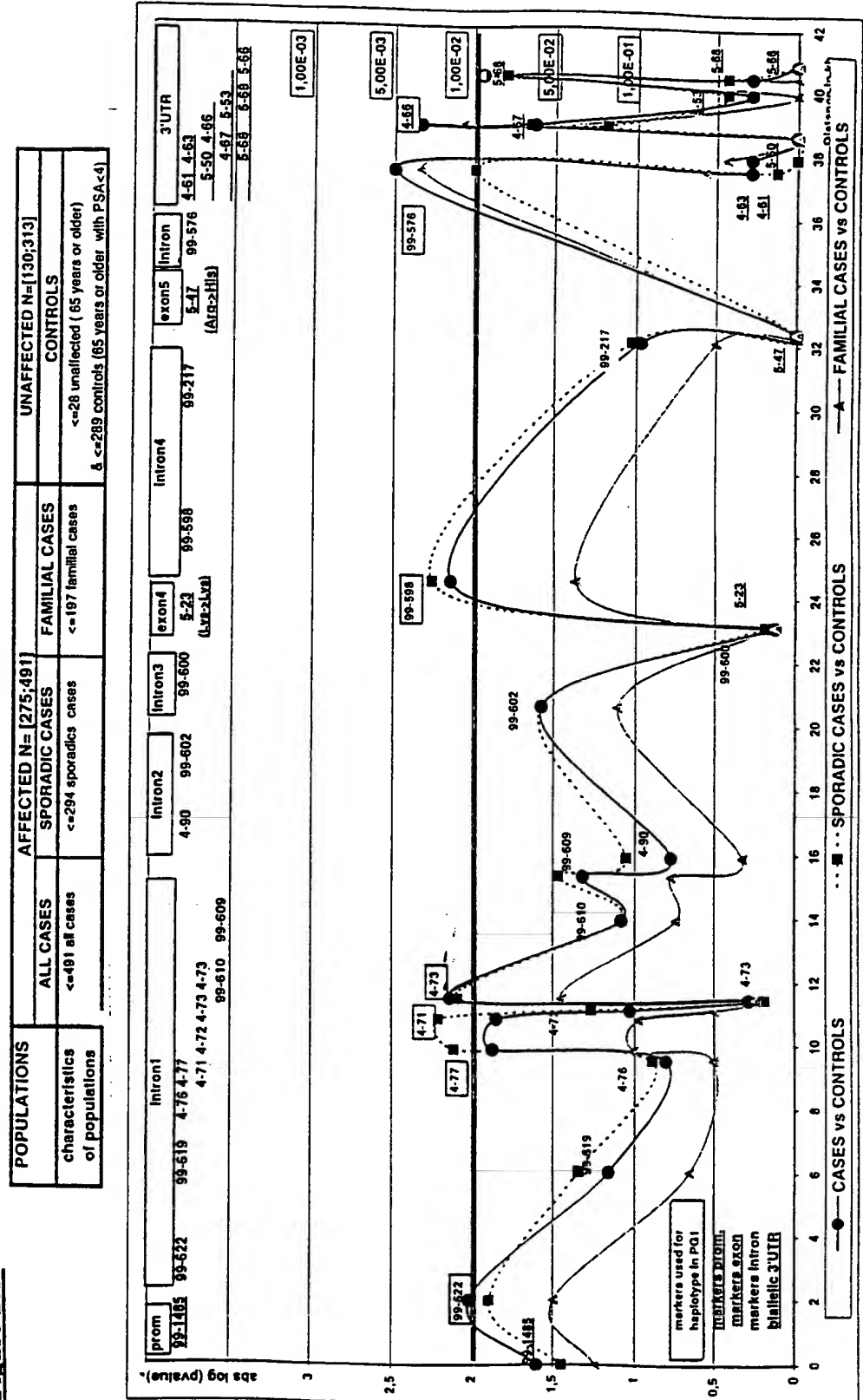


ASSOCIATION STUDIES PG1 (8p23)

Figure 17

POPULATIONS		AFFECTED N= [275;491]		UNAFFECTED N=[130;313]	
ALL CASES		SPORADIC CASES		FAMILIAL CASES	
characteristics of populations		<=284 sporadic cases		<=197 familial cases	
		<=491 all cases		<=28 unaffected (65 years or older)	
				& <=289 controls (65 years or older with PSA<4)	





20/24

Figure 18B

name of markers	PG1	Polym.	Freq(cas)	Freq(controls)	abs diff % ((freq(cas) - freq(controls)))	Odd Ratio	P-value	Freq(randoms)	Attributable Risk
99-1485/251...	prom	G/T	0,32	0,24	7,4	1,44	2,53E-02	0,29	17,58
99-622/95	in1	G/T	0,52	0,42	10,1	1,51	9,64E-03	ND \$	ND
99-619/141	in1	C/T	0,28	0,22	5,8	1,37	6,93E-02	ND	ND
4-76/222	in1	G/A	0,43	0,38	5	1,23	1,57E-01	0,41	13,15
4-77/151	in1	G/C	0,34	0,26	7,4	1,43	1,35E-02	0,31	18,16
4-71/233	in1	A/G	0,34	0,26	8,3	1,49	1,43E-02	0,28	18,64
4-72/127	in1	A/G	0,36	0,30	5,7	1,29	9,43E-02	0,31	13,25
4-73/134	in1	G/C	0,52	0,42	9,7	1,48	7,29E-03	0,52	26,76
99-610/250	in1	G/A	0,43	0,37	6,2	1,30	8,33E-02	ND	ND
99-609/225	in1	A/T	0,37	0,30	7	1,36	4,83E-02	ND	ND
4-90/283	in2	A/C	0,29	0,25	4,4	1,25	1,68E-01	0,28	9,32
99-602/258	in2	A/G	0,33	0,25	7,4	1,44	2,69E-02	ND	ND
99-600/492	in3	T/A	0,34	0,34	0,3	1,01	7,52E-01	ND	ND
99-598/130	in4	G/A	0,35	0,25	9,2	1,55	7,29E-03	ND	ND
99-217/277	in4	T/C	0,31	0,28	3,8	1,20	1,07E-01	0,28	8,46
99-576/421	in6	G/C	0,27	0,17	9,2	1,72	3,18E-03	0,24	18,40
4-61/269	3'UTR	G/A	0,01	0,00	0,3	1,76	0,527 \$	ND	ND
4-66/145	3'UTR	C/T	0,25	0,19	6,2	1,43	4,68E-03	0,24	13,16
4-67/40	3'UTR	T/C	0,25	0,20	4,9	1,33	2,39E-02	0,24	10,97

\$ Test Fisher - \$ ND: Not done - * disease associated allele / not associated allele

21/24

Figure 19A

HAPLOTYPE FREQUENCY ANALYSIS

POPULATIONS	AFFECTED	UNAFFECTED
sample sizes	CASES (n=491)	CONTROLS (n=317)
characteristics of populations	294 sporadic cases + 197 familial cases	28 unaffected (65 years or older) + 289 controls (65 years or older with PSA<4)

PG1 (8p23)		haplotype frequencies		Odd ratio	Chi-S	Pvalue
		cases	controls			
distance between mks						
size (cases vs controls)						
frequency % (cases / controls)						
abs diff freq. all (cases-controls)						
pvalue						
Hardy Weinberg						
Disequilibrium						
HAP 1 <43kb>	451 vs 297					
HAP 2 <17kb>	451 vs 296					
HAP 3 <117kb>	452 vs 299					
HAP 4 <100kb>	479 vs 302					
HAP 5 <60kb>	476 vs 300					
HAP 6 <160kb>	476 vs 303					
HAP 7 <160kb>	447 vs 297					
HAP 8 <60kb>	446 vs 294					
HAP 9 <117kb>	450 vs 296					
HAP 10 <160kb>	474 vs 300					
HAP 11 <160kb>	445 vs 294					
		0.116	0.067	1.83	9.85	(1.7e-03)
		0.243	0.183	1.43	7.49	(6.2e-03)
		0.182	0.130	1.49	7.18	(7.3e-03)
		0.217	0.188	1.20	1.88	(1.7e-01)
		0.155	0.132	1.20	1.54	(2.1e-01)
		0.373	0.346	1.12	1.16	(2.7e-01)
		0.095	0.042	2.39	14.62	(1.3e-04)
		0.117	0.065	1.93	11.33	(7.3e-04)
		0.178	0.125	1.53	7.80	(5.2e-03)
		0.114	0.089	1.32	2.44	(1.1e-01)
		0.095	0.032	3.18	21.59	(3.4e-06)

22/24

Figure 19-B

HAPLOTYPE FREQUENCY ANALYSIS

PG1 (8p23)

markers of haplotype Max	4-14/240	99-217/277 ln4	4-66/145 3'UTR	99-221/377
	C	T	C	A
distance between mks	<100kb>		<17kb>	
			<43kb>	

PG1	sample sizes	haplotype frequencies		odd ratio	chi-S	P value
		cases	controls			
cases vs controls	cases vs controls					
cases (<=65 years) vs controls	445 vs 294	0,095	0,032	3,18	21,59	3,40E-06
cases (>65 years) vs controls	171 vs 294	0,105	0,032	3,56	20,91	4,80E-06
	271 vs 294	0,079	0,032	2,60	12,13	4,80E-04
sporadic cases vs controls	266 vs 294	0,086	0,032	3,23	19,73	8,60E-06
sporadic cases (<=65 years) vs controls	85 vs 294	0,095	0,032	3,20	12,04	5,00E-04
sporadic cases (>65 years) vs controls	178 vs 294	0,085	0,032	2,82	12,75	3,50E-04
Informative sporadic cases vs controls	67 vs 294	0,062	0,032	2,00	2,70	9,40E-02
familial cases vs controls	179 vs 294	0,098	0,032	3,32	18,33	1,80E-05
familial cases (<=65 years) vs controls	86 vs 294	0,112	0,032	3,83	17,98	2,20E-05
familial cases (>65 years) vs controls	93 vs 294	0,075	0,032	2,48	6,59	1,00E-02
familial cases (>=3 cap) vs controls	78 vs 294	0,123	0,032	4,26	21,33	3,70E-06

HAPLOTYPE FREQUENCY ANALYSIS (PG1)

Markers in PG1									
size (cases vs controls)									
allelic frequency % (cases / controls)									
allelic frequency % (randoms)									
diff freq. all. % (cases-controls)									
pvalue (cases vs controls)									
Odd Ratio									
Attributable Risk %									
<div> <div>Hardy Weinberg Disequilibrium</div> <div> <div>cases</div> <div>controls</div> </div> </div>									
haplotype 1	2 MKS	339 vs 167							
haplotype 2	3 MKS	330 vs 122							
haplotype 3	4 MKS	312 vs 122							
haplotype 4	5 MKS	311 vs 121							
haplotype 5	6 MKS	309 vs 121							
haplotype 6	7 MKS	290 vs 99							

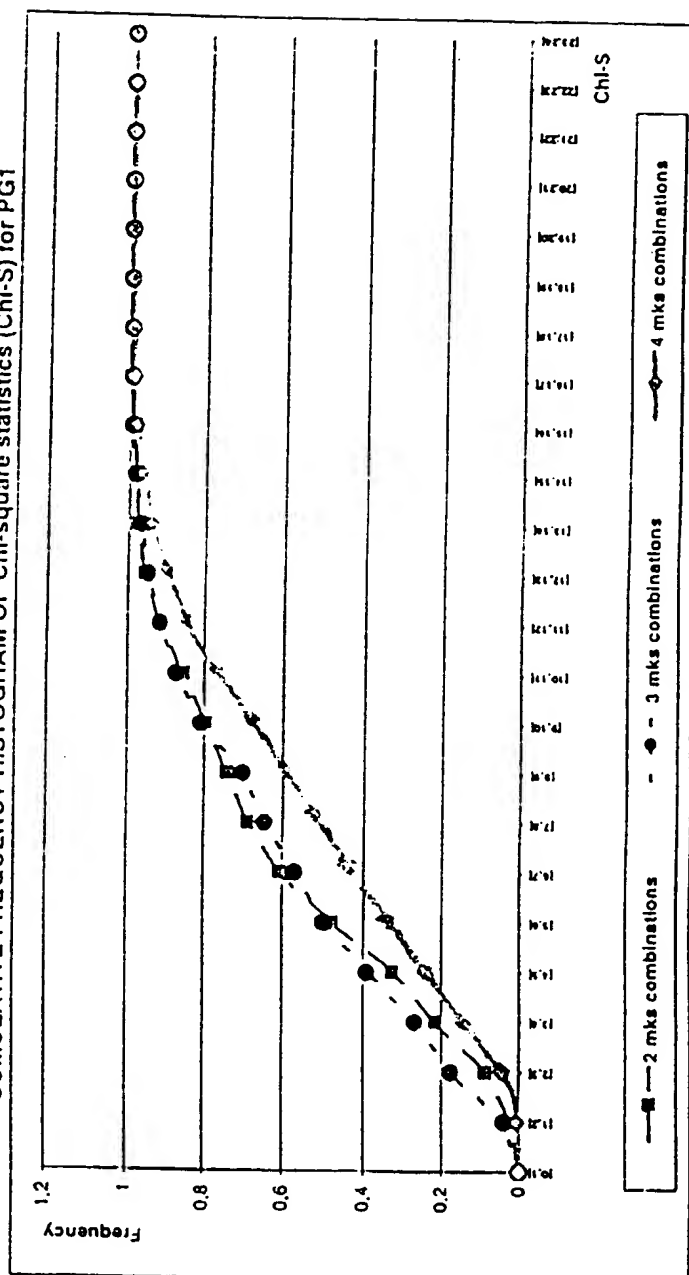
24/24

Figure 21

Comparison of Pvalue between nb of mks for haplotype
(19 mks of PG1)

GENE	# of markers	# of 2 mks combinations	# of 3 mks combinations	# of 4 mks combinations
PG1	19	171	969	3876

CUMULATIVE FREQUENCY HISTOGRAM OF Chi-square statistics (Chi-S) for PG1



<110> Cohen, Daniel
Blumenfeld, Marta
Chumakov, Ilya
Bougueleret, Lydie
<120> Prostate cancer gene
<130> PG1
<150> 08/996,806
<151> 1997-12-22
<150> 60/099,658
<151> 1998/09/09
<160> 578
<170> Patent.pm
<210> 1
<211> 56516
<212> DNA
<213> Homo sapiens
<220>
<221> promoter
<222> 1629..1870
<223> identification method Proscan
<221> misc_feature
<222> 1998..2000
<223> potential ATG
<221> exon
<222> 2001..2216
<223> exon1
<221> misc_feature
<222> 2031..2033
<223> ATG
<221> misc_feature
<222> 11694..14332
<223> Tyr Phos
<221> primer_bind
<222> 11930..11947
<223> upstream amplification primer 4-77 SEQ ID42
<221> allele
<222> 12057..12103
<223> polymorphic fragment 4-77 SEQ ID24
<221> primer_bind
<222> 12339..12358
<223> downstream amplification primer 4-77 SEQ ID51, complement
<221> primer_bind
<222> 13547..13564
<223> upstream amplification primer 4-73 SEQ ID64
<221> allele
<222> 13657..13703
<223> polymorphic fragment 4-73 SEQ ID58
<221> primer_bind
<222> 13962..13981
<223> downstream amplification primer 4-73 SEQ ID67, complement
<221> exon
<222> 18196..18265
<223> exon 2
<221> exon
<222> 23717..23832
<223> exon 3
<221> exon
<222> 25571..25660
<223> exon 4
<221> primer_bind
<222> 34216..34234

```

<223> upstream amplification primer 99-217 SEQ ID43
<221> allele
<222> 34469..34515
<223> polymorphic fragment 99-217 SEQ ID25
<221> primer_bind
<222> 34625..34645
<223> downstream amplification primer 99-217 SEQ ID52, complement
<221> exon
<222> 34669..34759
<223> exon 5
<221> exon
<222> 40688..40846
<223> exon 6
<221> exon
<222> 48070..48193
<223> exon 7
<221> exon
<222> 50182..54523
<223> exon 8
<221> primer_bind
<222> 51149..51168
<223> upstream amplification primer 4-65 SEQ ID65
<221> allele
<222> 51448..51494
<223> polymorphic fragment 4-65 SEQ ID59
<221> primer_bind
<222> 51482..51499
<223> downstream amplification primer 4-65 SEQ ID68, complement
<221> primer_bind
<222> 51596..51613
<223> upstream amplification primer 4-67 SEQ ID44
<221> allele
<222> 51612..51658
<223> polymorphic fragment 4-67 SEQ ID26
<221> primer_bind
<222> 51996..52015
<223> downstream amplification primer 4-67 SEQ ID53, complement
<221> polyA_signal
<222> 54445..54450
<223> AATAAA
<400> 1
gtggatctgt gactgttcgc aggaagagag gagcgggagc aggacagaca ataactgata 60
gtcaggagct gggtttggag ataaagaggg aacaagagaa agttaagttc tgtgttttca 120
tggcaaacat tgcacaaaag tttacaactt cgtgactaac agtaatctgg ggtgattcac 180
aacaaattta cacataaaca catatttact gactttatac acagcaatcc taacgtgaac 240
acagaacctg ctttatcttt tcgcacactg ttctagtgtg gagatgtctg gtctcagtta 300
aagaaagcat aaggagcatt agttgtgcac actgtccaca cccgtgactt ttttccacca 360
gtactaaacc tagtgcttct tacagtacag ggcaatgaca gccacagaaa gagagaagct 420
ccttttactg tgtaatgctt cctgctggcc ttcaaatact tgttacttga gagatctcca 480
ttcacctggc ttgttcccca aagggtcatca tctaccaatg atgttggtat ttgatgttaa 540
tcatgtataa agaaagtagc taccatcctg gccctgatta gaacttccca ctgaaatacc 600
gtcctgccta aaggtagcac aggtttccat tatggtggtg gtggggaggg ggcgggaata 660
tatatatata tatatatata tatatatatg gtaaagcatt cggcattctt ttaaagtaca 720
actatccttg aaaagggtta catattaaac catttttacc acagccaaag gggaggagaa 780
agatccaaaa gtctgtgga tctgctttaa catcaataaa acagttatcc acccttcgta 840
gcttttagtg aaggctacaa aagtatgctt tttatggatt acacatgtgc acgcaactac 900
tttaattact acagaaaaaa acgaggctcc ttattaaaaa aaaatcagaa acaagtccaa 960
cagactctga ggaaatgaag caagagtga ttctgaaaag gtctaataaa cagtatggaa 1020
atatccttgt gggattgttc ttcagctatg cataaacatg taattatcat cattactgtg 1080
atggggaaaa acacggacc taattctgaa acaccctggt agcgaagagac gggcaggagg 1140
ggctgctgcg cactcagagc ggaggctgag gaggcggcgt ccccttgcaa aggactggca 1200

```

ctattgaaaa	tccagttaag	tctctctact	gtgttgagag	gcattgattc	aagtacctgt	4920
gttactttcc	tgtgctgcca	aaacagatca	cctcaaaacta	agcgggcttaa	aataatagaa	4980
cttaagttct	cgtgattctg	gaggccagca	ctttgaaatc	aagggtgtagg	ctcaatttta	5040
ctccctctgg	aggccctagg	gggaatctgt	tcttgtgggt	ttcaacttct	ggtgactggg	5100
ggcatttcct	ggcttggggc	cccatcactt	caacctctgc	cttacagtc	ttgctgccac	5160
ctcttctgtc	tcacatctca	ctctcccttt	ctcttagaag	gatgcttgtc	attgggttta	5220
gagcccacct	ggatattccg	ggatgatctc	ttcatctcaa	gacccctaat	tataactgca	5280
aagagccctt	ttccaaataa	gaaaacattc	acaggtrcca	gggcttagga	tgtggacaca	5340
ttttttgagg	gggtgccctt	cattccccca	caacaatgaa	ctccatagtt	ctgcctattc	5400
agtattttgt	agttatttcg	tagtttaact	tgccattatt	cttttaggtat	ttacgtatta	5460
aagcattttg	gtctctgctt	tctttaacag	agaacctggg	ttctctgaat	aagtttacct	5520
actttcccat	aatcttttag	ttctttattt	acagatttac	cttcacatat	cccttaagta	5580
gaacatttga	ttactgtttt	tatttttcgga	acaaatctgc	attctgtata	ataaccaact	5640
tattcatatt	tcggtattct	tttaattctt	atctgattct	gaaattacca	tcttgtgatt	5700
atatatatat	atatatggaa	ataactgaaa	tcttgataaa	ttaaagggtga	tataacttct	5760
aagacaatta	attatgtatg	atgtgggtgaa	tatactgggtg	tttgggtttgt	ttgccactta	5820
aaagccctat	ctataggata	ggaagtaact	tgaatgtgga	atgcttagag	actcagagta	5880
agagggccgt	tatatatcct	tgagctggag	tttaaggaaa	acttatggga	aattaaaagg	5940
aaagtgtgag	tactgacaga	ggattgcgta	ggactcatga	aaaagggaatg	aagtttacct	6000
aaattctatc	atcgtgagtt	aacgtgaaac	tagattttatg	ttagtttata	gcctagaatt	6060
ctatcctagg	aatctagata	tatcctaaat	gttgagatag	ctgcataaac	aataactgta	6120
atcgttatga	taaataatga	caaatctttt	tagcatgttt	tgtgaagctg	ataaatgtta	6180
ataggatgtc	ttcaaatgtc	agaattcttt	ttcttttgc	tcttttttaa	aaaaattctt	6240
ttcccccat	cctatgcaat	acactgaaaa	ctgatcattg	aaatttttag	gccaaaaaat	6300
taatcaacac	gtaatagatt	ggggtttggg	tttttttag	tcagggtctt	cttctgtcac	6360
ccaggctctg	gtcgggtggc	accatcatgg	ctcattgcag	ccttgaatgc	ctgggttcaa	6420
gtgatectcc	ggagtagctg	ccgtgccatt	atttctagct	aattttttaa	agtttttgta	6480
gaaatggggg	ctttctgtgt	tgcccaggct	ggcttgaat	tcctggcctc	aggtgatcct	6540
tctgccttgg	cctcccaaag	tgctgggatt	acagggtgtga	gccaccatgc	ctagccctca	6600
ataaatattc	taattaccga	ttatctttgc	ttaaatcagt	tggtaacact	tggaaattac	6660
ttcagaatat	attttacatt	agtggctctg	actgctaatt	cccccttctc	caaatgctaa	6720
tgtaatataa	caataaaatg	cacagttctt	aagtttatat	aaaataaaca	ggttttcagt	6780
tgacctgctt	taagtgtaaa	atagtggtgaa	aaacacaaaga	aagaagataa	agaatttaaag	6840
attttgacat	ttctctaata	tgcccttaac	ttctccaagg	attcatactt	ttttttgtaa	6900
gacagaatct	cacactgttg	cccaaaccag	aggtgcagtg	gtgcagcttc	cactcactgc	6960
aacctctgcc	cccgggctca	agcgggtctc	ccacctcagc	ctcctgagta	gctgggacta	7020
cagggtacaca	gcaccatgcc	cagctaattt	tttttttttg	tatttttttag	tgggggttaga	7080
gacgagattt	tgccatattg	cccagctctg	ttttgagctc	ctgggctcaa	gtgatccgtc	7140
cttgatccac	catgcttagc	tgattcctac	tcttaactga	aacattgttc	caagtttctc	7200
agaacagtc	aaggcttttt	atctagagaa	catttataac	tggatctttc	tttgtgtagc	7260
actgattcat	caaactaatc	ctaaactcct	aatgagttaa	atttatattc	tgaatcttgc	7320
tgtaaaagca	gccattcatt	agaatgaaac	atgtttactt	agaattggag	aagggaagctt	7380
ataagtcac	tagtctactc	ccttttatga	cacttctaca	ttctttctgc	acttctgcca	7440
aaatgttgcc	cagcgtctgc	tctgatacct	atagtcttaa	caagaatatg	aatcatacct	7500
tgatctctta	attttactct	tctctgctta	tttgccattc	atgtgaagac	cttaaataga	7560
tcttaaatg	cttcttccac	tttagctgag	agtgacagga	ctgtgtaggt	gtgggtgtgt	7620
ttctgcat	gcttatttaa	gcaggataat	aaaaactttt	actataggaa	attaaacatt	7680
tcccaatcaa	atacaattcc	agtctaacac	aattaaattc	tggttaggga	actgcttaac	7740
ttactagact	tataggaaaa	tactaaaaaa	atgtaactag	aactctattt	ttacacttta	7800
taaataataa	cctctgtgaa	caaaccagtt	atttcagggt	gcatttgtgt	atagtttttt	7860
aatgcctgat	ttttctattt	taaaatcaca	gatgcaatta	tacattcaaa	cactgccaca	7920
atactttgag	aaagttaaag	tttcccttac	tcctacactg	cgtacacctt	tcctagggtac	7980
atcccagttt	gggtgtgaac	tttagatttc	ttccaagagc	ttttgagtaa	gtgtttgaat	8040
tgtgggaagg	ttcttttagt	aaatgaactt	cttacagatc	agtttttttag	tacagtagca	8100
cgaaatatac	ctgcatacct	atggggatag	ctctgtgcca	ttacgatgga	aggcacggga	8160
aaacagcact	ccgtatatac	ctagtttact	ttccctcttt	tgtatatgtt	tctgattttg	8220
tggagctgat	gcttctcaag	tggaaatcaga	agttaaactt	tcctttacta	ttttctcatt	8280
ttattatggt	ttcttaacta	gaggttgatg	ttagtgggtg	gaccattcaa	tagtaagtaa	8340
tgacttttca	gtaagggatc	tctagaaccc	agatccctta	attcctgcaa	tattcccggtg	8400
tgtacattgt	tccaggtgct	gtcctgggta	ccaagggata	caatgtttga	tagacaatgt	8460
acctgccatt	atggaggtca	cattctagtg	tgggaagaca	aacaataaca	agaaaaatgaa	8520

aatttactgt	gccatgccag	gttgttttagc	ctggtgggtg	agaggtaggg	gtttggaaaa	8580
tcttactgag	caagtgcacat	ttgtgtggag	ctctgtaaaa	gggccagctt	ggaaggtaaat	8640
gtagtcatcc	agggtgagaaa	tgatggttag	gggagtggaa	agagtggatg	ttaagattga	8700
aaagaattcc	aaatctatctt	tagtggttagc	tgatagggct	ttgtgattga	atgtggagga	8760
aaaagaagag	gggtgggttag	taacacactc	agtcgcagtt	agtgagtgtc	gctgtgtgca	8820
agtattgttc	tattatgtaa	ataattccat	ctttacaaag	taggcacat	tcttcctctt	8880
ttacagacaa	ggaaaaggga	acacccatgg	ttcacatctg	tagtagccta	gccaggagtt	8940
tcaggcactt	attttctgaa	gatgctctgc	ctggcaatgt	ggttatatgt	gttgaaatga	9000
gacccccctac	tttcaaggta	ttcatctagg	aaagacatga	actgccaat	acaatatagg	9060
ataacactga	aattagagac	gtgtttatta	actttgccat	acagaggtaa	agtaactctt	9120
taaagtaact	ctttgcttgg	gttagtggag	aaggctataa	aaattacttg	gagtttttac	9180
tttgaacatg	cgtaattaac	atgggaatgtt	taggggaaaag	aggttttcaa	ttgataacat	9240
aataaacatg	aggagtttga	agcatggcat	tcaaggtttt	ctaaattctg	ccccggttaa	9300
cttttccatt	cgttgggttc	attctagtct	agcttttctt	tctgggccc	ccctccccac	9360
attagaccgc	tctctcttgg	aattccaact	caagcccttg	cttttctcca	tctgtcatga	9420
tgttacccca	tctcattgtc	agggtaactt	ttatgtaata	ttaacatata	taatactgat	9480
ataacattag	catattttaa	tgtatggatc	atctctctctg	caacattgta	acctcttgga	9540
gatggcaata	atgggaagaa	tgacttgatt	ttactttttc	ttttaacaaa	aatgggtggag	9600
tagtctgggc	acgggtgtggc	tcatgcctgt	aatcccagca	ttttgggagg	ccaaggaggg	9660
tggatcactt	gaggtcaggc	attcgagacc	agtcctggcca	acattgtgaa	acccccatctc	9720
tacaaaaaaa	atacaaacac	ttactgggca	tggtggtgtg	tgctgttagt	cctagctact	9780
caggaggctg	agggtgggaa	atcacttgaa	catgggaggt	agaggctcca	gcttgggcga	9840
cagagtgaga	ccctgtctca	aaagaaaaaa	aaggtaaaag	ggccaggtgc	ggaggctcac	9900
gctggtaatc	caagcacttt	gggaggctga	ggcaatggat	cacctgaggt	cgggagttcg	9960
agatcagcct	gaccaacatg	gagaaacccc	ttctctacta	aaaatacaaa	attagccggg	10020
cggtgggtgg	cctgcctgta	atctaagcta	catgggaggg	tgaggcagga	gaatcacttg	10080
aaccaggag	acagaggttg	tggtgagcca	agatggcacc	attgcactcc	cgactgggca	10140
acaagagcga	aattccgtct	caaaacaac	aaacaaacaa	aacaaaacag	agagaaaagg	10200
cagagtactc	tagggaattc	tagtctgtgt	ttctgtggaa	atgtatatga	atctcacttt	10260
taagggatgg	agatttttga	atggcataac	tagttgataa	gttttgctct	aacagggtac	10320
ccaagtctag	tgagtcctgat	tcattctttc	cttaaataga	tgaaggagga	agaaacatga	10380
ctccaccctc	aagagttaagg	cagaatgagc	aaagtcaag	aagttaaaaa	agaattctca	10440
cgcagccagc	agtgagagaa	aaccttggtt	tagttgtgaa	tcaaaaccag	tactttttgt	10500
aattttttgag	cctatgcaat	tctccaaggt	tttatgttgt	ttcttctgtt	tctctgtagg	10560
caccagaaat	caaaacccca	aataagaaaag	tgttacttga	agatttttaga	gtactttatt	10620
gtgtataagt	gtaagtata	tttggaagac	gactttactg	cgctcctcca	gcttggcatg	10680
agaattccag	gggpcggaag	aaaggagggt	gatggtagct	ggaaaggaga	gtcatgttaa	10740
gtcccagcca	catatttaagt	gctaaccacc	tactgttaaa	aggtgtaatg	ttctagactg	10800
acaaaataca	tagtctctac	cgtaaagtaa	cacataat	agcagtgcag	aaagatgtca	10860
cttaaaagaa	aacttgataa	tatgctgaga	tagttcacaa	attaaagaaa	tgaacaaaaga	10920
actgaggaaa	taaaggagga	atacaactgt	gtccaaatga	atacttaact	gggtgggagc	10980
tggtgcata	gtaagcaggt	ggttcaccta	aaagtggat	gtaacgtagt	taacgccagc	11040
tcttggtgca	cttacatatt	gcattgcttc	cgggttaaat	ttgtgttcat	ataggaataa	11100
atttttttgtt	ggtttttta	tttactcctt	gtaattccgt	ggttgatatt	caaagtga	11160
aaaattacat	aagcttctaa	tatatgagaa	gtcttctcac	ttgacatttt	ttatttggaa	11220
tttttgcaga	gagtagtttt	gtcacagtca	aaagattttg	ggatcttgca	gtgagaaacc	11280
taggtgtaat	tcttatttct	ctgccattcc	gtatgtcatc	tggtattaagt	gtcaacttct	11340
cagtctcaag	attctctgtc	ttaaatggaa	tactttttgt	catgctattt	tgaagacaaa	11400
atgagataat	acgtgaaact	gcctagctca	gtgaatggta	catcatagat	actcagaaaa	11460
aacacaccct	ctaaaataag	aacagtagca	aaagacagga	tgtaaaataa	gggcagtagc	11520
aaaagacaca	tgcatgctga	gtgtatgaga	aagaactttg	tgcccttctt	gggtggcaca	11580
ggccatggca	gttccacagc	atgacgtggg	tgctgtgggt	ggtagagcag	acatgccgct	11640
ccccgtcact	gcctggcttt	gatgcttgct	ttcttcagct	gagaggacgc	agctgtgata	11700
tgaagggtctt	gtgtgtacag	tcgtgacctc	acattttcaa	tttcctgctg	gcagaaccca	11760
cagtctacaa	cgtacgagca	ccagagttga	cgtgagacag	acagcataca	gaggcttgta	11820
acatccttct	ggpaaacact	gtgtaagctt	tcagtgcgaa	taaacatgat	cagtggcaag	11880
ttctgttaga	tgtagtctgc	aagcatcctg	attttactgg	gcaagactat	gttgattttac	11940
aggcggtgca	tgattccatg	gatagccac	tactagtatt	ttcacaaatt	tcacaagaca	12000
ttcttactgg	aagattgccc	tgttcttatg	atactgctgc	ccttttagct	tcatttgctg	12060
ttcagactaa	acttggagac	tacagtcagt	cagagaactt	gctaggccac	ctctcaggtt	12120
attctttcat	tcptgatcat	cctcaaaatt	ttgaaaaaga	aattgtaaaa	attacatcag	12180

caacatatag	gcctatgtcc	ttgagaagca	gcagttaatt	acctaaacac	agcaagtacc	12240
ttagaactct	gtggagttga	attgcactat	gcaagggatc	aagtaacaat	aaaattatga	12300
ttggaatgat	gtcāagagga	attctgattt	ataacaggct	atgaatgagt	acctttccat	12360
ggtcgaagat	tgtaaaaatt	tgttttaagt	gcaaacagtt	ttttattcag	ctttgaaaat	12420
gacttgcata	aatctggaga	aagattatca	ggatttaata	tggtgaatta	tatggcatgt	12480
aaacatttgt	ggaaagcaag	tttgaacat	cacatattct	tctgttttga	cagaccactt	12540
ccaactagaa	agaatttttt	tgacatttat	tttaccattag	gttcaaaaatt	cctaattgcat	12600
ggtggggagaa	ctgaagttca	gttagttcag	tatggcaaaag	aaaaggcaca	taaagacaga	12660
ctacttgcat	gatcctcaag	taagccattg	acgtggaaat	taatagtttg	ggaagtagta	12720
ggcaggaatt	caatatctga	tgaagagatt	agaaacataa	agccttccat	cacaattccc	12780
acccggaaca	ggaatttcta	ctcatcaaaa	ttctgcattc	atacaagagg	gaacctgatt	12840
atgaccatct	tctgttggtc	atttggtaga	ttatgtgggt	cacacttctt	ccaaatattt	12900
gcaaatcaga	catcaccatt	atcagcacaa	gctaatagca	tcatctctga	atcatcata	12960
ttacaggaca	cccctggaga	tgggtagcct	ccagctttac	cacccaaaaca	agctaagaaa	13020
aactgtttga	accaaattca	ttattttcat	tttcaacaag	atctggaaga	tcatattaat	13080
gaaacgttga	tggtctatct	tctcttaaaa	aatctgctcc	taatgggtgt	attctacatg	13140
ataatcgtgt	tctaattcca	gtgaacctga	cgaaaatgga	aggtttggag	tcaatgcaaa	13200
gggggatatg	atcagaagat	gtctgtgac	gtgtcctgag	aagcaccagg	aacacctttg	13260
acctcagtg	ctctcgattg	aagagaagac	caagttgtat	tgatcagtgg	ttgggacttt	13320
acagaacaca	cccattgattg	gggtgtcctg	ctttttaaag	ccaactgtga	gagacattct	13380
ggggaactca	tgcttctagt	tctacctatg	ctgcatatga	tgtagtggaa	gaagtgtcag	13440
aaaatgagac	agacttccag	tacattcttg	agaaagcccc	actagatagt	gtccaccagg	13500
atgaccatgt	gctgtgggag	tcagtgatcc	agctaaccga	gggcttatcg	ctggaaacatt	13560
ctggacacaa	tttgatcaac	ttatcaaaaa	aaaacttgga	atgacaattt	ctgggtgccg	13620
attaccttag	aactcttgca	aaaatagata	gagatagttt	tccttatgat	gttacctggc	13680
ttatttttaa	aggtaatgaa	aactacatca	gtgtaattcc	agcatcataa	gtcagaacag	13740
tgcttgtaa	ggggcgcttac	cacacacttg	aacagatttt	tggcagatga	cttgggaaaca	13800
aggctcctcc	atgtttgtaa	tggtgaccac	acaagttgaa	tgtggcagag	ttaaatgacc	13860
ccaatattgg	ccagaaccca	caggaaagttc	atcctatgga	tgctaccaag	ccttctgcc	13920
ctgagaagaa	ggaagcactg	tctttatctt	caggaagatc	acactgctgt	ttaaccaaga	13980
gaaaaattag	agagtcacat	atcacgcaga	tccagtacag	agggtggcct	gacctgggag	14040
accctgatga	ttcagtgact	ttctggattt	tgcttttcat	atgcaaaaata	agagggttag	14100
caaggaaaaa	ccccttggtg	tttcttgcat	tgctggagtt	ggaagaacca	gcgttcttaa	14160
tactatggaa	acagccatgt	gtctcattga	tctcattgaa	tgcatgcagc	cagtttatct	14220
actagacatg	gtaagaacaa	tgagagagca	gtgagccgtg	atggtccaaa	cacctagtca	14280
ttacagtttt	gcgtgtgaag	tactattttg	aaagcttatg	aagaaggttt	tgctgaagaa	14340
agcaaaagga	aaaaaagaac	tttgtcatct	gttaggttcc	attttattga	tgataattgt	14400
gtttgtattg	attatttggc	aagtagctgt	ttgctatttt	gatcttattt	cagaagggca	14460
taataatttt	actattcaat	gaaacgtttt	aaacggggta	gaaaaagact	agtttttcta	14520
tgctttacag	cagaaatctt	ataatgatta	actggttaata	tatttcgttg	gcataaaaaa	14580
acatttaaaa	gttcaagtaa	ttataaacat	tgtaaatgtt	atatgtaatc	atattgaaat	14640
tgaatttctt	tatagctgta	cttctgtgta	atcaaaagact	ggggagagat	agactagcta	14700
gctctttctc	ttatccatta	atcacttaac	agagttttga	ataaaaagtt	ccatttcatg	14760
ggataagaat	aatgacaggt	taacctattt	tagttgggtta	ctatgttcta	ggtgttgtat	14820
gaagtagttt	acatagtttc	actgatttca	ctacaatccc	aggaggagta	gttactatta	14880
ttacactcat	tttacaggca	aagaaatagg	tttggagggg	ttgggtgttt	tgcccaagtt	14940
ctcatcgtaa	aatgacagat	gaggattcaa	attcaagtct	taattgaagt	ccattacttt	15000
agaacctacc	tcttagtggc	tcttatgtta	cagtataagg	gagagcagac	tgttccttta	15060
cccttgtagg	gtagctaggg	cttgtgaatt	aagagactga	ttaacaggag	aagaggcata	15120
cacattttat	tgacgttagt	atttttcat	gcacagggaa	ggagggtttt	atttttattt	15180
ttatttttat	ctttatttta	aagagacagg	ggtcttgctg	tggtgccagg	gctggactca	15240
aactcctgaa	gccaaagcat	tcttctgctt	gagattcctg	agtagcaggg	actataggtg	15300
tgctcctctg	tgcttggtca	aagaagggtg	ttgtatgtga	tttttaacaa	aggctgataa	15360
atttgaaga	agtactagt	caaaggagaa	gaggatttca	gctcccaggg	gtggtaaaat	15420
gtgggaagat	gactaggaaa	tgtatagtaa	taagggttgc	tatgcaggtt	tattttgcca	15480
gtttctggtc	tcctaataag	ggacagggaa	acacctttac	agatggaaat	tcatatcacc	15540
tttccacagg	gaaatttatg	tcctgcctta	ggcagtttag	ggaagggcag	agaattcttc	15600
ctgtatctgc	tggtgtctcag	gtgccttcag	ctcaaaaata	tccttatgcc	aaagtagcat	15660
atttgggtgt	ggcatattct	ctgatctctt	tcaacagcat	catctatact	taacaacagc	15720
aaaagttttt	tttaaaaaat	catgtttcaa	gatttgcatt	tggaagacaa	atggacatga	15780
ttgagataaa	tgaagaatat	atatttttta	acaaagaatg	ctgtatatatt	atgtctctgt	15840

gacattgtgt	tatggaggct	aagggtgttaa	gcattgtgatt	acttttagatg	ccgtatgact	15900
acctgttttt	aagattaaaa	aagaatcaat	aggcagttta	tatgcatggg	agcaagttaa	15960
aaacaacaca	gatgtgatga	aggcgagggtg	aaactgggtcc	gcatctaatt	caggccttct	16020
cctgaaagcc	agtgtgtgca	agataaataa	gtttgtttga	cgaaagcaga	ataactagtt	16080
tgctcctttgt	gatgaagata	gttattcaga	aatcattttt	attggctacc	tctgaattaa	16140
taaatgaaaa	gagaaatttt	tttttctgta	ggggatgtct	gatgagttct	taaaaagtg	16200
atgaacctga	aattatcatg	aacaagcaat	tataatgaac	ttaaaattac	ttaaagagtt	16260
atgaaaaaca	aaaagaaaag	ccgtatgttt	tcttgtgcct	tattttgaag	tgacaaatta	16320
tttgcagggt	acattttag	acggaaactaa	tgtgatttaa	aaaatgagta	ctagatttac	16380
agaatgatgc	ctttaaaaag	tcactgggtgc	actttaatta	ttttatttat	gtttattctg	16440
aaactacctt	tattttgaaa	atgagggtata	gctttgccta	ctggtgacaa	aagtgtaaat	16500
aattcagtaa	acatctgtta	aaaaccagct	tggtgctagg	ctcttgggg	agaaaaactga	16560
tcaggccatt	gaggagctca	tagtccctaa	ggggctgggg	acttgtcatt	aggtgtgcag	16620
tgtgttctgg	atgctcctga	aggagtgtgg	gcagggtgcgc	accaccatgc	ctggctaata	16680
tttttataat	tatgtagaga	cagggtctgg	ctgtgctgcc	catgctgggt	ttgaacttct	16740
gggcttaaga	gatcttccct	ccctgcccct	accgaccctg	cccgccact	ccaccctcgc	16800
ctcccaaaag	cactgggatt	gcaggcatgg	gccactatgc	ctgggctgtg	caaaaacttt	16860
aaatcagtgc	atpctcaatg	gtcttgatgc	aattctggct	tggttgtaag	agaatgggga	16920
tttactcaca	agccacgatg	tcacttttaa	ctctgaacag	atcaagctat	tggtattact	16980
catttatgtc	atpgataaac	tttatgaata	aaaactcatt	gtgcaaatat	ttaaaactac	17040
tacatacata	gcactgtgca	gtttctaagg	aaagttaagg	aaacctttgt	cacatccctg	17100
gcttccagaa	ctttatgtta	tctaagtga	tttgtctgca	aagttgttgg	gttaattgcc	17160
cctttctttt	ttctcttttt	aagatattaa	taaatagttg	catgaccaaa	agataatcct	17220
tatggacaag	atagatctaa	aaagccttag	ctaatttata	atcttgcata	atccatgatg	17280
acaagatgca	gaaacaaaaa	tgcccagaat	aaaaacttag	caccattagc	agccatttcc	17340
ttttaagtct	ttacaagtat	actcccagtt	tcttgaaaaa	tttattctaa	aatatgtaag	17400
acacacaaaa	cagcagaagg	actaatacag	gtacatcgaa	cacctgtgtg	cctaccgccc	17460
agtttaaaaa	taaactggaa	tgatgtttct	ctcatactta	cagaataaag	ttttaatctt	17520
tagcatggaa	ttcaaaagac	ttctgccatt	ccagttcaga	gccacccttc	tggtctcctt	17580
gctcctcagc	cgcgacactg	cccatgtacc	caacaggcct	ccagggttac	tgcttccatt	17640
cgttcttatt	ctcatgaaca	ttttccttca	tctcatctgc	cagaatccta	cctaataata	17700
ctcctgctct	gcagtttaca	gttcttttaa	attaaaaaag	gttgtgtacc	ctttagtgtc	17760
ctgaaaaaag	aaaaaacaaa	tttaaaacct	taaaaaagga	ccatattttc	atagtatttg	17820
cgttatgtct	cattacagtt	cctgtggaca	tgtctgtctc	ttttactaga	ttgattgtgg	17880
gctctttgaa	ggaagatata	tcttatgaac	agtgttttat	atattgttat	caatcaatga	17940
atgcttgcta	tatttttctc	atgaggatat	tgattattct	attttaattt	attaccnnnn	18000
nnntgtacta	tacataactg	ctttctgtac	ctgagctatt	tatgatctct	gaggctcctg	18060
tgagaaatct	aatttttgg	aatcatggat	ggaaatattc	acaacatcat	tcgtcagttt	18120
cttcacattg	tcttcccttg	tatattacag	atgtttttaa	atatcaaagt	aatgtttttt	18180
tgttttatct	tttagatatt	gctatatgga	gatttgccaa	aaaaataaag	aaatataata	18240
tatttagcaa	atcatcaaag	cacagggttg	tatttcattt	gcatgaaacc	taggtttttc	18300
tacagatggc	acatgggcat	tcaaaaatacc	gttcttatat	ttaaaatgaag	tggttttttt	18360
aaaacagcaa	tttctgtg	agatattaca	cctgttcttg	tattttttgtg	attttacttt	18420
ttggaaagtc	agaaacttga	aagctatgaa	ttttcctaaa	cttaccctct	ccctctgttg	18480
gatgtaagta	agctatcttc	ttacttgctt	gctttgtttt	tcctttgtgt	agctctttta	18540
agagtgtatt	caftcttttt	gtaagtgatg	tttctagaag	tagcattggg	gggtcgaagt	18600
gtgtatacat	tttacatttt	tgattgctaa	gctgcagaaa	agctgtattg	gtatgtaagt	18660
actcgtttcc	ttactatgct	cgtcatttct	agtgtctgct	cttcccttcc	ttcttcaaat	18720
gggtttgggt	taattctagt	tgctactgtt	ccatcagagg	aattgcagag	aactgggtctt	18780
caaaacagtg	cagtatatac	tttaggtgaa	gatacttcta	aaaacctttg	tatttttgagg	18840
taattctaga	gttccaagaa	tttgcaaaaa	gagtacattg	tcagcaatat	ttttcccaat	18900
gggtgacatct	taatataact	gtagcacagt	agcagaatca	ggaaattgtc	attgggtaag	18960
gtacttttta	attctccaaa	taattcagcc	ctccaaaaaa	atcccacttc	ttatgttttc	19020
aaacctgtag	ctacttttga	tgcgacttcc	ctaaattgca	tttttattac	tttaaaaaat	19080
ataataccta	gaagctcaaa	gctggaaaca	gcctgatcaa	tatagtactc	ttaagctaaa	19140
aacaacctga	tcaatatagt	actcttaggg	aaatcactta	tgctgtgggc	tttttttaaa	19200
ttttcttctc	gtcagctgtc	tcttcatgat	tttgtgggtt	ttattactgc	ttataccata	19260
gatgaggtat	agaaagtaaa	agaagttaaa	atgcattttt	ctcaatttag	tgaattaatg	19320
attacattca	gatttatagg	acaaggtttg	aagctancaa	ggggttgata	ggaatcttga	19380
tgtatctgag	tattttcccc	aactttatta	catgactggg	tcagactatt	ttatctaat	19440
acatttcact	cttggcagaa	atagcaaaac	agtcaaccaa	tggtcaatgc	tgctgagAAC	19500

tctggcctgt	gcagacatat	tggctgtttt	acttctaata	ccattctgct	tttccctgtcc	19560
tgctgctgat	ggatgtttct	tccaggtttt	aaatatcaaa	caaaagggat	ctgtggggccc	19620
agtacaggga	atggctcttg	atagatttga	ttttcctgca	tttcctttat	tttgatccag	19680
tgtaatttc	atgtagagtt	gtctgtttaa	caggattctc	ttaaaattcc	ttcttcagtt	19740
tacctgccag	ctttcttttg	tccaggtttc	agtatgaact	ccactcgatt	aatagagctc	19800
tctagtagtg	acttggtggag	tgggttctct	gaacatttct	ggaagtgttg	ctgatatgta	19860
taatatgat	cactagtact	gttaatttgt	gtgcttacta	catgttggct	tttatatgta	19920
ttccttcaga	ttaaggactt	ctagaaaaca	tccatgaaaa	aacagattaa	aaaaaacaat	19980
tctgcatgta	tttgggacta	gaaggtacta	tgggaaggat	aatcttcata	ctcagagctc	20040
actgacctga	atttcattta	tcagtttaga	gaaccacttc	cccttcctct	caccctacct	20100
ccgagtgcct	gtgactttgt	atcaccgctc	tggcaccaca	tcctcatccc	agcaggattt	20160
gggaaggctg	ctttttgaaa	gcctttttaa	attctgttaag	ttgagaaaaa	actaggggaa	20220
tgatttttaa	ttcttttaga	attacaggct	ttagttagta	tatgacagag	ccctttccta	20280
gaaaaatgtg	catataaaaa	tttgcatgta	gttttagggt	ttcagagacc	cctaaagcct	20340
atccatagac	gtggttcatt	gtctgattgt	gttttaggtac	ccttctaaaa	cccttttgag	20400
atgttaggaa	tcacaacaga	gtatctctga	aaatgtaatt	agcggaaaga	acatttcaaa	20460
gactgttgtt	ctgcttagac	ttcttagttt	gtcttctgcc	aggcttgccg	gaataaatga	20520
gtttcctggc	ctgatactca	aaagaattga	cattttaaatt	agtctctctc	ttcccttggt	20580
ttcgcttgac	acatccttgt	ctctacattc	tgtctctgtc	tctgttagct	tatttctctc	20640
tcgagtcagc	aggatatagt	ggctgttatt	tcttcccctt	atccttcaac	gatctacttt	20700
tgacaacact	ttgccttttt	ttttttgaga	tggagtttca	ctcttggttg	ccaggctggg	20760
tgtaatgggt	caatctcagc	tcactgcaac	ctttgcctcc	cgggttcaag	ccattttcct	20820
gcctcagcct	cccagtagc	tgggattaca	gacatgcacc	accacgcctg	gctaattttg	20880
tattttcagt	agagatgggg	tttcaccatg	ttggctcaggc	tggtcttgaa	ctcctgacct	20940
caggtagtct	gcctgcctcg	gcctcccata	gtgcagggat	tacaggcggtg	agccactgtg	21000
ccctgcctgc	tatttgcctt	tttaatctca	tgaatgttc	tcttttcttg	gctgaagtgt	21060
cacttttctt	gttgaacagc	atgcgtgggtg	agtagaatgt	tataaaaagg	gatggacttt	21120
ggagtttagag	agacccagggt	tcctgttctgg	cattgcagaa	atgctgttct	gcaataggct	21180
gtgtgtcagt	gggcaaatga	cttatctctc	agagccttat	tggttaagggt	tgagttagag	21240
ctcctttcag	gcaccttaca	gaggctgtct	cctaactctg	gtagcgtacc	tggctcatag	21300
atggcattta	aaagtgtgtg	tgatgacagt	catagctcac	cattagcata	gcgctggatc	21360
catggcaggg	aagcgttgca	catgcagtat	ctcttggaact	acacagggcc	ctcatgaatt	21420
aggaactgct	gtttcatgag	gatagggatg	aggaatttag	acttgctgcc	cctcactgcc	21480
ttccactcct	ctcctccaag	ttaatgggaa	ctatgactct	gctttggctt	gattgccatg	21540
gaagattctc	acacagccaa	atttattgct	atcttagtta	aattatgcc	gaacacaaaa	21600
tatgaagtta	ttgtcaaaat	aatataatct	cagctgtaac	tgagatagtc	agaaactgtc	21660
tgtaactctga	tgctctatct	gaaaggtagc	tgagaataaa	caagaaataa	agagaattca	21720
gtagcaata	ttggtgacac	aaagctttta	tattttgact	agttaagcta	gttcttaaat	21780
gtttccacta	aaatattcaa	gtttaagggtc	atagcccagg	gcagcttatt	atgaacatga	21840
tgtatttttg	aaatcttaca	ctttctctta	aaagttcttg	ggaggggcat	gtgaggccat	21900
aataaacca	taaaaccatt	tgttttataa	taaaaccctt	ttttaaaatt	cttccaataa	21960
aaaaaattat	tcaggaaaaa	aatgctaaac	ctggttttta	actttgtacg	ccaactatat	22020
ttccaagatg	tgctgtagcc	tggttaacct	acagaacct	acagaattag	ttctcagaat	22080
ttattgtctg	cttacttttg	catttggtac	aggtataaca	gggtcgatta	tatgggttct	22140
aagacatgac	tagaaagaaa	tatgtttatc	agttattatt	tcttccattc	aaattagaag	22200
gggctaggga	gagggcttca	acaggaattt	atatacttta	gagaaaagtg	atcattgata	22260
gcccataagt	atagatatct	caacccaata	acacagggtg	tgtctgtctc	tgggatcata	22320
cactgtaggg	gagaatcttt	gcaagcaaca	ttctacttat	agggagccat	aacaaaagtt	22380
tcataatgtat	aataattata	agtcttaagt	catcaagaaa	aagtttaact	gtgaatgata	22440
atccctgatt	aaaaagagag	atgtataata	atggataaga	gatttttctt	ggtttaattt	22500
tagtattaaa	atggctaaat	cttctttggg	atattctgac	tagtatgggtg	cattgtctaa	22560
tagatttccc	atagctgaga	gctaactcat	ttgtaatctg	tggaaaaactg	tcctctttgg	22620
ctaaaacttt	attgtaattc	ctctaaatcc	tcagctttta	ttttctacag	actttttttt	22680
ttttttaaca	tttcccttct	ctgactcact	ccttttggtc	tcattttcat	ggcctgagaa	22740
catgggtgat	gatagaatta	ttcttttcac	agattaacag	ttttcttttc	gagtatcggt	22800
gagctcatgt	gtgtattaac	tagagaagtc	tcccttacat	ttcattttta	tgttttcttt	22860
ctcatcagga	gatagtttgt	agccatttac	tttcaaatcc	aagtttctgc	ggttcttaag	22920
acctgtatca	tttgtctcct	gaatttcact	tcatttctct	tttaaaccat	gtcctctgtt	22980
tcccatcttc	tgcacccact	ttgccacttc	ctgtttgttt	aattggcaag	ggccactctc	23040
tgtgttggaa	attttttctt	tttgaagctc	caactaacaa	cttctaggaa	gttttttatt	23100
gctactgtta	tcaattcata	ccatcttacc	cttgtttttg	caaccctttg	ttaataacat	23160

atttatttaa	ctatagttat	tagcagttctg	agatcatttt	acttggttac	ataaggagca	23220
catatatcta	cccagcatca	ttgtaaggca	tgtgagacct	ttgtttgatt	gctgtcctaa	23280
cctagtagcg	agtcctaaaa	actcattagt	agaagatgaa	gtgtccttgc	cttttgccta	23340
acatatatat	acacactgaa	tatttagtgg	caattcatag	ttgcatttgg	ccattttttg	23400
tttataattt	ccpctttctc	attaaaaaaa	ctttgttttc	tagacttttag	gatttagaga	23460
agctcatttt	gtcccataca	catgctgctg	ttggattatt	taggtatttt	gtgactgtat	23520
tttatctttg	aaataaaaaa	cctttcaaga	aatgcaaaaa	aaaaaagctc	aaaaaacaga	23580
aaatgtatat	tttttaataa	tctcagatag	atttaaagaa	attttaaaaa	tcctaatacat	23640
agtacttttg	aagcccatcc	atagtacaac	ctgtgaagag	cctcatgtac	gcgctaactg	23700
ggtcctgtct	ctgcagttga	ctggattgtt	gctgacatct	tggccatcag	gcagaatgcg	23760
ctaggacatg	tggctacgt	gctgaaagaa	gggttaaaaa	ggctgccatt	gtatgggtgt	23820
tactttgtct	aggtaacttg	tttccatgct	tttctctcta	tatatgtagt	ttataaattt	23880
tttttttttt	ttttggagac	agtctcactt	tattgctcag	gctgagtgca	gtggtgtgaa	23940
cacagctcac	tgcagccttg	acctctgggg	ctcaagtga	cctcctgcct	ctgcctccca	24000
agtagttggg	accgtagtgc	ccaccatcat	gccgggctaa	atttctattt	ttttgtagag	24060
atgggggtct	cgctgtgttg	cccaggctgg	tcttggaactc	aagcaatctg	cctgtctcag	24120
cctaccaaaa	tgtctggatta	taggtgtgaa	ctgccatacc	caaccctata	aaaatgttat	24180
atttttaaaa	taaacaatat	acttcatgtg	aatgtatggt	tttttaaatg	ggtttaatat	24240
tttattctca	gttgaaagtaa	ttttgtttgg	catttttagt	gggtgtgtatt	tatatagctc	24300
tgattatcca	tatgctggtt	tccttcagca	tctgtgggga	ttgggttttag	aaccaccaca	24360
gataccaaaa	tctaaggtgt	tcaagacctt	catatagaat	gggatagtat	ttgcatataa	24420
cctgtgcatc	acttttaaat	atctctagat	tacttataat	atctaataca	ttataaatgc	24480
catgtaaatg	gttggttatc	tttatttttt	atttgtatta	ttttaattgt	tatatatttt	24540
ttaattttta	ttgtttcaca	tatttttgat	ctgtgatttg	ttgaatctgc	agatgtggaa	24600
ctcatggatg	tgaagggcca	gctgcagtaa	aatgaaagag	caaaaaatgc	aatgtacaaa	24660
gttcaaacaa	ataggaaatt	taaaggcata	gaatttgata	ggcaattaca	ttaaactgtt	24720
gataacagta	attagtgtatc	tgtatgatat	taaaaaaaaa	aagcaaaactg	tatatataaa	24780
acttactttc	tccagttctg	gaggctagac	atccaagatc	aagggtgtga	cagggttagt	24840
ttctcccaag	gcctctctcc	caggcttgca	gacagcatcc	ttcttctgtt	gtcctcaggt	24900
ggtttttttc	cctgtgcccc	agcaccctct	gcactgcttc	ctcttcttag	aaggactagt	24960
tacactggat	gactaatcct	tctacagaga	ctgctaagg	cccactctga	ggcccttttt	25020
taaccttaat	taccacctct	aagtcctctt	ctctgaatac	agtcacagt	ggaactatta	25080
gggctttagt	agactgattt	gggggaacac	acttctgtcc	gtaacagtgc	cacataaata	25140
ttcttagcag	gattgatttt	ttaaaatccc	taaagatcgt	gagttatgac	atgttaagga	25200
cgcttttttag	tgactctgta	ataagtggtg	ggaagaattg	ggagttaaat	ccatctgatg	25260
gatcagggtt	tttattttta	aaaatgtgta	tttaagaaag	aaagcatttt	catttttaact	25320
gccaacaaaa	ctaaacttca	tgtgttttcc	aatacagtgt	cacatgcagt	ttttttgaat	25380
tatgttgaga	caaggcaatt	ttcagctaaa	tgttcttttag	aagctaattg	ttgaagatat	25440
taaatataga	ttaaattctg	aaatgtagtt	ttcattctgt	actttttgca	agagaagttg	25500
cctttttgat	gactctggcc	aattgttatt	ttaaaagtta	atgctctttc	tcccgatttg	25560
attgtggcag	catggaggaa	tctatgtaaa	gcgcagtgcc	aaatttaacg	agaaagagat	25620
gcgaaacaag	ttgcagagct	acgtggacgc	aggaactcca	gtaagagcct	accctgtttt	25680
atttttctta	ccagctctca	gtttctaaat	ttaagaatta	aattaaaaac	taagaattgt	25740
tttgacaatg	tattttccca	tgtgttaatta	ctaattcagg	gttatgctga	ggtaacagaa	25800
acctcttatg	tacaggtagg	cagggttttc	agccatcaga	aagattgctg	taaacaacta	25860
ggctccttgc	tggctcagtg	accttaaaag	ggaataaaaa	gagcatttgg	tgctgttcag	25920
agtcctataa	tagaactaac	tgcattttta	cctgacattt	aagctagtgt	acaagctcat	25980
cttacttctt	gtcttcttta	gtatcagatt	tggtttttaga	agcagcaact	gttttctgtt	26040
agtgcaaat	ttgaatgtct	tacatgtaca	gaaaaaacca	aaaaggatga	atctctacaa	26100
atgttaaatc	attcagtgta	aataatattt	tataaaactt	tattccacaa	aagtggggag	26160
agttcaatct	gctttgtata	gaatgctgat	tgctgccaaa	ggcttttccc	ctgggttccc	26220
ccggagacaa	agcaccatga	tcaccggggc	gacttgggct	ttctctttca	gtacatgaca	26280
tgtgctcaga	agcttagctc	gtgtgcacag	gctttcccct	tcctttctgg	ctcccctccc	26340
ctgtcttccc	tcctctcctc	ttgcccctcc	ctcaccaggg	gtcctgggca	gcagctggag	26400
ctcatgggtga	aggaagaatt	cttcatggtc	agctggcgaa	gtgctgtgtg	tgagcattgt	26460
ttattcacat	gcctcttcta	gggtgtttta	cattagaaca	ttgcatctgt	tttgggcatg	26520
tggttgggtga	cagaagcaga	atggaaatgag	atgaacagt	accctttatc	ctgttatagc	26580
taacccttga	gaaccaagct	tgggtgtctc	aaagggtctg	tttagtctga	aacagtgtgg	26640
tgaatttggg	cagaattgtg	gtcattgcat	gtaggtctcc	aaaagacaga	ataagttggt	26700
aatatgggtt	atcgactttt	tacaaaaaaa	atttaaaaat	catgaattta	taccttaaaa	26760
tgtecatccc	acttctctcc	cagctgtcca	gtcaccctcag	caatggatga	ctgctgtgga	26820

gttccttctg	tgtcctgctg	tgggcattgt	atatatgaag	caaatagaaga	tagctgcctt	26880
ttgggtgatg	ttggcatcct	atgcacagtg	gtcccttgct	tttttgcccc	catgaatata	26940
gctgccagtg	gcgctagggc	tgaaaaaatc	agctctttac	acttgctcatg	tgtcttggtt	27000
atgtggctgc	cttcgtgagt	ttcttcttgt	ttttggtttg	cagcagttta	agtatcatat	27060
atctgagtg	catttaaaaa	tttttacctg	gattggctct	ctgagcttgg	atctatgatt	27120
tggtgtctgt	tattaatttt	ggaaatttct	ttgtctctat	ttccttaaat	attattccta	27180
ccccagctct	tcttctccag	ttatgtttgt	gttggttcat	ttctcgctgt	tctttagttc	27240
ttagatgcat	tattcgtttt	ttgttggttt	ttttttaaat	tttttttttt	acgccccctc	27300
ccttttttct	ttttgtgtta	cattttggat	aatttctggt	gaccacacct	tgagttcatg	27360
gattcctcct	ttggctgtgt	tgagctact	ggtgagccag	tttaaggcac	tcttcacctc	27420
tgctactgcg	tgtttcatte	ctcacatttc	cctttgacct	tgtttcatag	tttccatctc	27480
tgtgctagtg	tatctatctg	atcataaagc	ttagtcacgt	tttccagttg	aacctttatc	27540
attttattat	acttgcaagt	ctcttaaat	ccttgcttga	taattccaac	atctgggcca	27600
tatctgagtc	tgcaaatttt	gattacttta	tctcttcaga	ttgtgcttta	ttgtgcttct	27660
gtcactactc	ctaagatttt	gcctaacgct	gggccttttt	tgtaagacag	gagaaatgga	27720
ggcaagtgtg	cttgataacct	ggaaatggat	agacttgctc	ttctgcttgg	cttttagtgt	27780
tgaggagtg	agtcagttcca	ctgaggaggt	gcactgcatt	tggtttttgc	tcatgtgctt	27840
tttctcacag	cttcaggttt	ctgtagaact	cattactttg	ttttaggttt	ggggatgtcc	27900
tcccgcctaga	gcttttcttc	agtgtctatt	tcacactcag	cgttttcaca	tagcaccttg	27960
gagtggctct	cttctttatg	cctttcccca	ctatacttct	tggtactttg	ttactgaact	28020
ctcgctagtt	tggtggtaga	aggagaggga	agggaagtgt	cttttcattc	ttaggggaga	28080
tctcaggggt	ggagccttct	ctgatcctgc	cttgccttctg	gctgtaagtc	tgtgccaggt	28140
atgtattcct	gcctttacta	agagtttttc	cctgttctct	tcaccagacc	tcatcgagta	28200
ttcatccgtg	ccpcatgggt	agcagggttt	tgttgccctt	gttcatcagt	ttcaggctgc	28260
tgttccatag	gaaaggtaga	aagaaggatg	tgggctgggc	cctgagccct	tcccacaggg	28320
ctgcttttcc	ctcccacaag	cctacatcca	gtcttccctg	accgagtggt	gttttctttt	28380
ttctttgtct	tgtgagtaca	caggaggtct	gtgggtcgag	cctgtgaaat	gtgctgcatt	28440
ctccttgtgt	ctgtagccca	gggggttcgtc	tgttccactg	gctcatactt	ggctttctgc	28500
aaaattgata	aaatttttag	ctaaattctt	tttactggta	tctgttacat	tggcccccac	28560
ctaaacaacc	acttgcatct	tgtttctcct	ttgagttttc	catctttcct	tgactttttg	28620
ggttagttgg	ttgccttgca	accttgcagc	tctctgaagg	gtctaagaaa	agtcatgaat	28680
ctacagcttg	tcagtggtgt	tgttgttgta	gggttggcag	tagtattcct	tcagcattct	28740
acatacttaa	tggagccgcg	ctcccatttt	tggttaataa	atttcaaaac	ttggaacaat	28800
gttagattta	caaaaacgtc	agaaagaaca	gagtgcttct	gtttattcct	tatatagctt	28860
tttttttttt	tttttttttt	gagttggagt	ctcggctctgt	caccagggtc	ggagtgcagt	28920
ggcacgatac	tggctcactg	caacctctgc	ctcacgggtt	caagcaatct	cctgcctcag	28980
cctcctgagt	agctgggatt	acaggcgtgc	accgccatgc	cgggctaatt	ttgtattttt	29040
tagtagagac	agggtttcac	catgttggcc	aggctgggtc	cgaactcctg	acctcttgat	29100
ccgcccgcct	cggcccccca	cagtgtctgg	attataggtg	tgagccacca	cgcccagcct	29160
tcttcatcta	gctttaacat	ctaattgtga	catcttacat	aacatgggat	atatttgtca	29220
aaactaagaa	ataaacattg	gtaccacact	attaattgta	ctacagattt	ttattcagac	29280
tttaccaggt	tttccactaa	tgtccttttt	ctgttctaaa	atacaatcca	gaatagatac	29340
aatccattc	aacttcagtg	ttttaaat	ttgtttttca	ttatatgaag	tgctgtgtgg	29400
tttttgcata	atctgttatt	ttggttttta	tcttcaagct	tgtccttgg	tctttaagtg	29460
ataaaggcat	aatttaaaag	gtgtgttggg	ttatttcagt	gcctaaagtc	ttgtctgagt	29520
cacttggttt	ctgctgttct	tgcttatggt	actttctttc	cctgtttgct	ttgttatctt	29580
cctttgctgc	tggctgtggt	tggttaagtt	atttgtggaa	atcagttgaa	gcctcaggtg	29640
ggagtgtctt	tctcgggaga	acatttctac	ctgttttagc	tgggccccct	aaggctcctc	29700
tagcgtgggc	cccacccaaa	cgagattctg	agttgaaggt	gaactgagcc	attcaggcag	29760
tgagccagg	gttgcagatg	cacgtgagac	ctgctcacct	ctcatttact	ttcacccctga	29820
gagtagagcc	tttgggtgtt	cgttcacttg	tctgattctc	tcttcacagt	tctattagaa	29880
gggtccatggg	tttgggtttc	tgtgcccttc	atcttatgag	tcttgtaaat	caaagttctg	29940
ttttatgctt	acttctgctt	tactgtgttt	gcttaatttc	agtcttaaca	tcttgccaac	30000
tcttgggtac	ttttaaaata	atgttatatc	cagcttttta	agttgttttc	agtaggaagg	30060
ttgattcaaa	taacctagtc	tggttatggg	ctacgagaat	agcctccctg	ttttttgtgg	30120
gcaaaattcc	agpccttttat	gttccctagc	cagtgtggat	aacagactgg	caggttcaag	30180
aggccgtgct	gagcagcttt	cactgtaagg	tcactgtccc	aggtcggttc	tctaagaatc	30240
tggtgggttg	tttcatttct	taatatgtac	gccctgtgag	agcggatata	tcttgcctcag	30300
gttcttatga	ttcttttgtt	tctgaagggt	aattaaagtaa	gtgacatggt	agaatatgtt	30360
aagtcaactt	tcgttgggct	tactagttct	catgaatcta	ttccatgatt	gtatcagttc	30420
ttattcagta	ttagtattta	agaaatgcag	aattttgttt	caaaaaatat	atttgtatta	30480

taagttgtga	agaaatacat	ctccataatt	attgctggga	caatacagta	ttttcttaag	30540
gaacttattg	gttgtggatg	caaatgaagc	atatttgtga	taaaaataac	taatagaagt	30600
cattttgtta	gactatgagc	tagtaaaaact	tatggcacaa	acatggagac	ttaacacttt	30660
ttcttccagc	tttcaactaa	gttccctttc	agataggagg	cagcctgggtg	gataagagta	30720
ttgggtttga	aatttagattc	aggtttaaat	cccagatcct	ctgtttaatc	tttattttat	30780
ttcaggtaga	ttttctggat	aacttgctat	agcttatacg	tcagtacttg	ccacttcaat	30840
tttatgttat	ggagagacgg	cttctttcct	taaacctcac	gaaccaacct	ctgctagctt	30900
ctaagttttt	tcttgccact	tctttacctc	tctcagcctt	cagagaatta	aagggagtta	30960
gggccttgct	ctggattagg	atttgcttta	agggagtgtt	gtggctgggtt	tgatgtttta	31020
tctagagcac	tcaaactttc	tccatattcag	caataaggct	gttttgcttt	ctaactcttc	31080
atgtgttcag	tgaagtagca	cttttaattc	tctttaagaa	cttttctctt	gcattccgca	31140
cttggtgtgt	tagtggaaag	gacctagctt	ttgacctacc	ttggctttca	acataccttc	31200
ctcactaagc	catttctagc	tattgatgta	aagtggagag	catgcaactc	ttcctttcac	31260
tggaacgctt	agcagccatt	gtaggggttat	taattggcct	aatttcaata	ttgttggttc	31320
tcaggggaata	gggaaaccca	aggggctggta	gagagaaaga	gagacaggag	aacaggccat	31380
cattggagca	gtcagaacac	acacgacatt	tatcaattaa	atattgtcatc	ttatattgggt	31440
gcaattcatg	gcacccccaa	acaattacaa	tagtaacatc	agagatcaca	gatcacataa	31500
acagatataa	taptatgaaa	tattgtgaga	ttaccgaaat	atgacacaga	gagctgaggt	31560
gagcacatac	tgttggaata	atggcaccaa	tagacttgct	cgatgcaggg	ttgtcataaa	31620
ccttcaatgg	gaaaaaaatg	caatttccgt	gaagctcagt	aaagcgaagc	atgataaaat	31680
gagatgagcc	tgtcactcct	aagaatgttc	ctgtacaagt	tttttgcttc	tgttacttac	31740
cttttcttat	ttgtgaatag	tatctttttt	gagtacgtgt	gtttttttat	ttttatacat	31800
ttatatgtat	cttttgaaga	acatactttt	aagcttaatt	tattgatttt	ttttctctca	31860
taattttccac	tttttgtatc	ctattttaaga	agtccttgcc	aaacttaagg	ttgtctaagat	31920
tttctccttt	gttttctttc	ggaaatttta	gagttttgct	tttactattta	gttctaggat	31980
ttattttataa	ttaatgtttt	catatgggtg	aagatcgaag	ttcatatttt	tttaatatag	32040
gtaaccatca	ctatagaaaa	gattattttc	ccccaatgtt	tgaataaagt	agactgaata	32100
tagatgggtc	tgttatccct	agatcaatgg	agcatttggt	ctgttatatt	gatctatata	32160
tatatatcct	tatgccaata	ccatactgtc	ttaataatgc	ttgcttttga	gtaagttttt	32220
aaatagtgtg	gttgtcttct	aaatttggtc	tttcttttca	aagttgtttt	ggctattttt	32280
ggtttttttgc	atttctgtgt	gaattataga	attagctcga	caatttctac	ccaaagtttg	32340
tgggctttttc	attttgattg	tattgaagat	atagatgaat	ttgggaagaa	ttgatataac	32400
aggattgaat	cttttgattc	atgaacgtag	cctgcatttg	tttacttagg	tcttctttat	32460
ttatctcagt	gtgttttgta	gttttaatgta	cagatttgca	catcttttgc	cagatatatc	32520
cctaagaatt	tcagtttttg	atactattgt	agatgacatt	taaaaaaatt	tcaagttttt	32580
gtttgttgac	ctaggcatat	atttgacttt	ttaataatac	aaccttgcta	aacttattta	32640
tcattctagta	acttacaaaa	tatatctcct	aggatttcct	acataaacia	tcattgtcatt	32700
gttttagaaa	taacagtttt	actttgtcct	ttttaatctt	gatggctttt	atttcttttt	32760
cttgctaaat	tttctggcta	gacctcctag	tacagccttg	actagaactg	gtgtgaggga	32820
aatccttttc	atattctcca	tcttttagga	aaagcactca	ttctttttatc	cattcttttag	32880
ttcctagccc	cattgcccct	cctaaatttt	ttctcatcat	tttctttcat	cacaccttgt	32940
tctttttctt	tgcaatcata	tcattgatag	taacgacatg	tttttattta	tctgtttaat	33000
gtatttcttt	tcptcacttg	tccatgaagg	gaaggacctt	atgtgttggt	atcctttgtg	33060
cagtttcttg	aacataataa	gtatataaga	aatagtttct	gaattagctg	tgaatgaatt	33120
catgccttcc	tgtgtgtctg	caatgttctt	ttaaattaaa	catctaagac	agcaataaat	33180
accacatgag	ttattaacct	gagaaataat	cgttttattt	ataaatgact	gagttgaaag	33240
ctgatagccc	acbgtaattg	ctttcatggc	tttgaatata	aaccttactg	ttacaaaaca	33300
cattttcatg	aaaatgaatg	tgtggtgttt	ggaactagct	ttaatgtttg	tcttctgttt	33360
tttctcttca	gttgctataa	tataataagg	aattttgtat	gttttttcta	attgtaccca	33420
cttttctaca	ttttcttaac	agatctgggtg	aatcttcatt	attaaatata	attatacata	33480
taaattattg	tttaataata	atattaatta	ttaaaaataa	tataaattat	taaatataaa	33540
gatacatata	atattatctg	ttaatctcta	agttaggtgt	gggttctgaa	gactattata	33600
tgaatgaaca	aaaagcttgc	atatttgcgt	ggaagctgaa	agtacgaaat	ttttagatac	33660
cattatacca	gtatctaaag	aaaaaattca	gtaccacata	ggttttttaag	taggagctgt	33720
atgatcatag	gtcatccaga	tgaaggaagg	cttctgtacc	agacgtacag	aggtagacag	33780
tgttgtctga	gtactgtctg	agatctggca	agaatgaatc	caataaacgt	agttttctcc	33840
catgagctcc	tgtcttgttt	cctgtattct	gtttgtattt	gaaaagattt	gggtgtgcata	33900
acttattttt	gtcttttggc	tgtcaatcaa	agttatttagt	gtagttttttg	taactcagtt	33960
ctcaagctag	gagtttttgc	tgtataattt	taatgtttct	gttttttactt	tcctaagcag	34020
ataagcgtaa	aaacttagac	taattgatta	cttattaaac	gtccagcttg	atattcttct	34080
ttatattatt	ttagtttcag	tttatataac	aaatgaggtt	tcttataaat	aaaatttaaa	34140

atgcactaaa	ggagctgtgt	gaaataggaa	ttctgtgtga	agcttttgaa	tgtgaacatt	34200
tagaacgttt	cacatgggtg	gaatttacta	tatgattttc	atcaaatgag	gtacttttta	34260
gtgtttgtac	ttaacgatac	tgattttctaa	aatttgtatt	tctaaaaatg	acgtattaca	34320
ggatctgaaa	gggcaaaaaac	tcattgaggg	tttgtatgag	tcagcgtttc	atggcctatt	34380
tttaattagt	gaattattag	catataatta	gaaatgtttt	tagatttctc	atggctgacc	34440
taccaatgaa	tgtagcactg	catttaaaat	atagtccacg	ttatgttcat	acttaattgt	34500
tgcattttgt	ttgccccctc	tgaacggaag	gtcacatgta	aataaatata	cattttctcc	34560
tactgttaga	aatactctgt	tagcattagt	aggttttagct	tttttaggtt	aacaataaca	34620
aaaacaaagc	tcacacaaaa	taaaccaaat	ttgctctatg	tcccacagat	gtatcttgtg	34680
atctttccag	aaggatcaag	gtataatcca	gagcaaacaa	aagtcctttc	agctagtctg	34740
gcattttgtg	cccaacgtgg	taagtaaaaa	tttgagtgtt	tgaacaaata	atcttcaaag	34800
ataataacat	ttttagtttt	tcttcctgga	aaagataact	ttgttttaca	gttgaaggaa	34860
tgaatgtatt	cattccctga	attagtgtac	atattatctc	ttaggaaatg	aagtttcttc	34920
tcctttaattc	actttcatgc	tattattaca	tatatctgag	aaattaagtt	gaagtgcttg	34980
ttacgataca	tattcttgtg	ccatggattt	atttaaaatc	tatctaagta	catgattatg	35040
tagatggaa	ctttttctac	agtgtatggg	ttatatgtaa	tggagcttct	gttttgtaag	35100
atgacagacc	taagttggag	tccaaactcg	tacttttatt	agctgtatgg	ttgcaacttg	35160
gaagtttgtg	aatgttgctg	agcttgcttc	ttcatctctt	aaaagaacat	agctcttata	35220
agtagatcta	aatctgtgtg	aggattagat	tagaaaatat	gtcaagtttc	tattggagaa	35280
gttacacaaa	gttggtccac	agtgtctgga	agctgttaat	gtcttcaaca	atggtaattg	35340
tcttaatatc	catatttttag	aaaattgaat	aattgggtaca	ccaataagct	atgcaattta	35400
accaaatgg	gaagtataca	gaaaacagtg	gctatgctat	gttcttagag	gtgtctttga	35460
agcttgactg	tgatttagtg	tgtgatctcc	atatgttgat	agtcactcac	tgagcaataa	35520
ccttgttggt	gacattacag	cagggcctat	gacagtgtcg	tctaattgaa	ctttctgcaa	35580
taattggtaaa	gttcttcate	tgttctgtcc	agtgtgtctg	ctcctaccaa	tgtgtttttt	35640
gagcattcaa	catgtgacta	gtgcatgaaa	ctaattttta	atcttattta	atcttagttt	35700
aattaaaaat	aagggggagt	ttttacaagg	tgcttacaag	agcagatatt	tcattaggtat	35760
atgacatcat	ttgtaacagt	acttttaaaa	aatggccagt	tgtttttaaa	cacatgtcct	35820
attaagtaag	gagtgtttca	gaataggagg	gttcagtgtg	tctcccatc	tgccagctct	35880
cttttgactt	tcattgtctc	ctctgtctaa	tagacatgac	gttctgtcat	ttcagttgct	35940
cttttgcaat	gccattgtct	cttttgccct	tttcacattt	attaaacaga	acaaaacaaa	36000
aaccactctc	gaatctgtag	tctacctttg	ttgttaagcac	tttttccagt	actcactctg	36060
ccctcaattt	gttttgggtc	gatttgaaat	tctctcccta	gacttctgtg	gggtgttctt	36120
ccattatcct	cccaactctc	tggtgattac	ttcctagcct	cctttccagc	ctctttctgc	36180
ttcatttctc	cctgtcatcat	gtgttatttc	cagtgtcagg	ttttgtgtgt	tgattaatat	36240
cactttttgt	ttctcatggg	ggccttccct	taaatccatg	gcttttagcca	tcgtttccct	36300
gactgtgat	gactcgcaaa	agcttccctc	cctccatgtc	tctctgccta	actctgggac	36360
catttgtaca	attgtccatt	agagagcttc	gcttgactgg	cccaaaaagga	tgcttcaaac	36420
tcagcatatt	gaagatagaa	tttatccttc	catgcataca	ctcatatttc	ttgtcttggt	36480
aactccatca	ttcagttttt	ttgcctaagt	tttattcaca	aaaagaacaa	attgatagca	36540
gttgcatacc	tcttatagga	aacttagaca	tggagggaaga	agctgttcag	atgggttctt	36600
gcagaagtgc	aggcactgtg	gtaatattta	aacttttctc	agctgttcga	aggggtttgt	36660
tttaactaat	tttccttaga	cttgttttag	gtatttggct	ttctaattgt	tataagggat	36720
gtggaattaa	atgtatctta	atctgccacc	tggaccatt	aaagtaagcc	cctatgggtg	36780
tttttttttt	taattgccat	ggttaaaacc	atagttgcta	gcgaagggtg	catacttaag	36840
ctttttgaac	tctcttaaaa	gaaaacagaa	atttaatgat	gtgtctataa	tggcaaacca	36900
gatacctaga	atttccatgt	tattcatagg	gtgaataaca	ctggcgattg	tagagatttg	36960
agagtctctt	caaaacagga	gaacaaaggg	aataagctac	aaagcaattt	ttttctttgt	37020
agacttaact	gaataaaaaat	tatttttatg	tctcaaacat	catatgaaca	aatttagttg	37080
gcaaattggca	agctaataat	attttataat	ataggatatt	aatatactta	atattacaaa	37140
agtgcttcat	aattagaaaa	gacataaact	agaaaaatgg	gaaaagggca	tgaataagaa	37200
attcaagaga	tacaaatgac	ccacacactt	gaacaaatgt	ttattcttct	tcataatcaa	37260
agaagttagaa	attaaatgaa	tactttgaag	ccaacttctg	agaaagcata	gcaaacaga	37320
aagctagtgc	tcagctttgt	gtggtaacgg	cactctcgct	cttaagaagg	tgtgtttgct	37380
ccctgtggct	gctctcaggc	agggccacaa	acttggtggc	ttaaaacacc	acagatttct	37440
tctctttacat	ttgagaagtc	tgaatgggtg	cttactcagc	tgaatcaag	gtgttggcag	37500
ggctgcagtc	ctttgtggag	gcttgggggg	atcttgttct	cctgtacggg	gtcctgtgct	37560
tgggttcgggg	tctgtgtgct	ggcttgggat	cctgtgcttg	gttcgaggtc	ctgtgctggg	37620
tcagtgctc	tgcttttacc	accttgaagt	tcactctgga	atggcactgg	ctcggccaca	37680
ccatatagct	gactctgggt	ctccctcttc	ctcactcgct	ctaaacctgt	gtttttggct	37740
gattttcaat	ctctctttcc	ttggcccttc	tgcagcttgc	agggccttct	gcagctcttg	37800

tctgccccag	ccccggggtc	tgcccatccc	agtgtctggg	tggtctgttc	ctgcccctgcc	37860
tttccctcagc	ccttgaggcaac	cctgtttgtt	ttctcccttc	cttagcagtg	gagaacatcg	37920
taagatcaat	gctgactgcc	ttctgcagcc	aagccaggcc	atttcatttc	agccgagcca	37980
agtctgtgtg	gagcagttct	tttatttttc	tccttttgac	tacctcatgg	ttttcacgga	38040
ttttgtttct	cttcacattc	aaggattttt	tgctttcaga	aagttatatt	tctctggaaa	38100
gagtgacccc	aatatccctt	ttgatttcaa	aatcttaatg	tggagtctct	tgacttggat	38160
ttctttggaa	gaaactgctg	aagctgccat	gtctaagaag	aaaactttgg	agaaaaattt	38220
tcctcttaga	catggcaacg	tcaacagttt	ctaagctctt	gattccgtct	accctgtctc	38280
catcgttgcc	tcagtcatct	gccttacttc	tctgcagggg	ttctccccag	cttgcaaatg	38340
tactccaatt	ctgaaataac	taagtctata	gctgtgcaaa	gagaagtctg	ggccccttgc	38400
ttctctgtgt	ttgactccat	ccactctcca	gaaatgaatc	ccacttctca	cttaaccact	38460
gacctccaaa	gcacgtatc	atttgtgtca	gttgtcatat	ttgttaactt	tcacataact	38520
tttgacatta	tttataccct	tataaccagg	aaataatttt	aactttattg	tagaaaaata	38580
caatggagta	taatttttct	tgttgaagat	aaatatcacc	tcctctctct	ttaaacatct	38640
cttccctttg	ttttgttatt	acattgggtt	ccccctttt	tttatttctt	gggttgcgt	38700
attccctgtt	attattttta	cctttttttt	tttaagtgtg	atgtttccgg	agtctgtatt	38760
tcttgccctt	tcattctctg	ccctttatta	ttctcagcca	ctgccattac	ttcagttatc	38820
cattcccatg	gtttccacat	gcttagcttc	ggttgattct	tgccatttta	cagaccatat	38880
ttccaactac	ttctagaatg	ttttgttctt	tcagcctcag	tatgcccaat	ttgaactcat	38940
gttctctctc	ccccctcttt	cttccctctt	tccttgcgtc	tctctccctt	ccttcttttc	39000
tttccctccc	tccttttctt	ccttccctca	ctcgttctct	cttgettgtt	tgctttctct	39060
ccctctctct	ttttctttct	gcnnnnnnnn	nnnatctctc	tcctcccttc	tttctctctt	39120
ctccccact	ccccaaactt	caggctaaag	cagtcctcct	gagtagttag	gactacagac	39180
atacacgtgc	caccgcgccc	agctccgtgt	tctctttgtt	tccttgcttc	ctgctcttcc	39240
acttatcttt	gcattggcagg	tgggtgcacg	caggcatgct	ctgcatgtct	tcctcttggc	39300
cattccctct	ctagttatgg	tgtggcttta	tctacgcgtt	ctggagcaga	agcctagtca	39360
caaagctatt	tttttaaaac	attcatgata	attcatttcc	ttttatgttt	taaaaaact	39420
agctttctgt	ctttatttcc	ttactaaact	acttggatgc	cagtaattag	ttgttttagt	39480
gaacaccaca	gagtgatatt	ttgaaacttt	ggacttcata	aagtggatg	agctccagta	39540
gcaaagaagg	aagtgttaac	tagtttaact	gacaaataaa	tgcttcccag	cttgggtgtc	39600
gattgagatt	ttgtgtgcaa	gtttgtgaat	caatttaact	gcccctgccc	tggggactaa	39660
agtcagatac	gtgcttgtgg	gaatctttgt	ctttcccaca	ccacccttga	ttttaaaacc	39720
tccttgtgtg	gacagtccca	ccatgtaata	gctgttcttc	cttactcagc	tactttccct	39780
ccagagaggc	cagtagaaaa	tctagactag	ttttttatag	tctattttca	tgtcacttat	39840
tgagagctac	tgttttctgt	taaatgttca	gtaaatattt	taatcaagg	aaagggaggc	39900
aataggaagg	agagaagaac	aaatccttaa	ccctagtagg	aacctaatga	atgggatttg	39960
ttctggataa	ttgcagtagt	ccccagcta	aagaaccttt	taaaaaatag	tcagatatac	40020
ccaagaggat	tgaatctgta	tgttcataca	aaagcttgtt	cacctgcagc	cttcatatgc	40080
aatcctatg	aatgttcata	gcagcattat	tcataatagc	caaagtatgg	atgcaaccaca	40140
aatgtccatg	aagcaattaa	taggtaaaca	aaatgtgatc	tgttcacaca	gtggaatact	40200
aactattcag	ccataaaaaag	gaatgaagca	ctgagtcctg	cagccacaca	gatgaacctc	40260
agatccatgc	tgagcgaaag	aagccagaaa	caggaggcca	tgtgctgtgt	gactgtattt	40320
ctaggaaatc	ttgagtcacc	atgggcaaga	tgctatcacc	tttgttcagt	ggccagaagc	40380
gagggcacta	atattttacc	ttgccggggt	ctactagatt	gaagcgtttc	cgctaggcca	40440
taaaacttcca	acacgggtgac	ttgtacatgt	agatatttga	tcaatatata	gcaaatgaat	40500
attgatttaa	acagaaaaag	gcaagtgaag	gtgctttcta	aacttagagc	cctaaaataa	40560
tgaggttgtg	gaattaatag	attctgttgt	gtgtgtttga	gggaatttaa	aaataattta	40620
gatgttaaac	agtatattgt	ggagggtgtt	tgtaaactaat	taatgacggc	actgaattga	40680
cttctaggcc	ttgcagattt	aaaacatgtg	ctaaccaccac	gaataaaggc	aactcacgtt	40740
gctttttagt	gcattgaagaa	ttattttagat	gcaatttatg	atgttacggg	ggtttatgaa	40800
gggaaagacg	atgggaggga	gcgaagagag	tcaccgacca	tgacgggtaa	gtgtgttcac	40860
gcacctgaaa	tgccctgtaca	cggtatatac	agtgacacatg	tttatgtaga	attcagtttt	40920
acaaagtagg	ttaagtgtac	tttttctctc	cattacattt	accgggtata	ttttcaaga	40980
tgttattaag	atgtaacagt	ggagatttca	ttagtctctc	aaagtgtggg	atttcttggc	41040
tgtcgtgtga	gtcctgtgga	ctcaccattt	atcattaatc	cagcctcttt	ctactcaaa	41100
ttcacactta	aaaggaaagc	tctgtaaaag	ggagggaagc	gtgaagaagg	agcacgcctg	41160
gcagtactga	gtgcacgtta	ttagtacgtg	ctgccctttt	gctgtatttt	tcgtaaaaata	41220
tttattaaat	ttgggtgtca	ttgtgacaag	aagaaatgca	gttaagtgtg	accttttttt	41280
ttccccaaac	atgtttagggt	ttagaacact	ttgagctatt	gtcagatata	accagaaaaa	41340
aatagaattt	taagttagca	ggataactta	gttaaaactaa	ccaaacatag	tgttagctgt	41400
tagagaaatg	taaacatgga	aataggcaaa	cagggaagtg	tgtggagttt	ctgttctctt	41460

ttcaaaatat	ctgttttgagc	tgggggttgag	agagaacact	aggcttcatg	gggtttttttt	41520
gttttttcgtt	ttttgtttttg	agacaagagt	ttcgctctgt	cgcccaggct	ggagtgacgt	41580
ggcgcaatct	tggctcactg	caacctccgc	ctcccacgtt	cacacgattc	tcctgcctta	41640
gcctcctgag	tagctggaac	tacatgcgtg	tgccaccatg	catgactaat	atttgtattt	41700
ttagtagata	tgggatttca	ccttggtggc	caggctgggc	tcaaaactct	tacctcaggt	41760
gatccacgca	cctcggcctc	ccaaatgagc	tttggttttt	tacctcatca	gctgtttggg	41820
gttgagccac	tatgtatgtc	agtgtgcttg	tatcagtagg	atctactgag	ggcagatgtt	41880
caaaatatga	gcctccagca	cgttttcat	ggaaaccctc	acctgaagca	ttcgtctgaa	41940
gttgatgtgc	cttggaattt	ttatagagta	atatttttaa	ctacaacaaa	acatttataa	42000
aagtagacat	tattaaagca	ttcagaagtg	agcaaggata	gaaattattc	tgccccacct	42060
tacacgtagg	ccttctagac	gtagtactgt	gcaccgttac	attatctaac	actgtctgtg	42120
tgctcatctt	ggatgttagg	gatttttcca	aagttcagtg	agattatagt	tgctcaaatga	42180
ttagtctgtt	aaataaatga	aagatgaggg	tcactcaggt	tttaaaagaa	aagctctttg	42240
actgaaagag	agagcagctg	tctactgcag	aaagttaggg	agggaggctg	gaggagtgag	42300
gcccgagggc	tagctagtat	aaaaattggt	tatggtcgaa	ggaaaaaaa	atgtaacata	42360
tttatatctg	aaagatgatt	gttctcataa	ttgtatataa	cacagagtaa	ttgtaaagta	42420
gaaaactaag	gtgtttttca	ttttagatgt	aaatgtttag	aatatgtaat	gcattacgtt	42480
aaaaattaaa	actgtacgaa	atgcacagtg	aaacgtcttc	cttgctttcc	accctgctac	42540
ctggccttcc	cttctccttc	ctagcgataa	ccagttttct	taatttggtg	tgcgttgtat	42600
gtgcaaat	aagtatatct	tcttattcta	ccatccctcc	cttcttacag	aaaagtggca	42660
tattaatatt	tttctctttt	aaactatcga	aggagtact	tacctatttt	tgcatctcaa	42720
aacagacagt	tcatcaagat	tgtcgttggt	ttattaaaca	tagtttaaga	ttaaacaagt	42780
gtttataacc	aatgaaaaac	agatagactc	cccataataa	ccttgtttaa	atgctgctac	42840
ttttatcatg	tcccctcctg	tctaagaacc	ccttggttca	gcagagctca	tgggtaaggc	42900
cagcctctgt	tgccctgccat	cggagggaatg	cgctccagcc	gtgatctctg	ccttgccctc	42960
gttctctcct	gtgctgtgcc	gtgaagcctc	ggccgtgggtg	aagctggctg	actgagtcct	43020
cctgcacccc	atgcatattc	agtagttgaa	ggctttgtgt	ggccaatcct	gctttccaca	43080
ggaaaccacc	ctctcttttg	ttgcctcat	ccaaggctac	tgcttctcca	gagtgacagg	43140
ggcaccttt	cccagcatag	cactgtgcct	tctcctgccc	ctgctctctg	agctactgct	43200
tggcactgat	ggcgtgtgtt	acagtgtggt	cacttagcac	agggctctgc	ctttctctct	43260
tcccagccgc	atcataagtg	ccttgaggaa	gccaaaacct	tctgtgagtt	gcattgcctg	43320
ggttccaacc	tcccactgcc	ctgcttatcc	tctgctacat	gtgagctgac	tgtggctttg	43380
gggtgggtcac	tgccctatgtg	tattcattac	aaattgtctc	cttttgaaa	attgaccttt	43440
ctgacttacc	cagataccat	aaagaaaata	aaatcttatt	acttcagtca	aggataaagt	43500
atttctgaat	taaaggaaaa	atacaccaga	gtaaaatcaa	gactgaaaga	caaactggga	43560
aattatttgc	aacctagatc	atagaaaagg	ggctatttcc	ttcttgcgta	aagtgcactt	43620
acaaaattgat	aagaagatga	ctgataacta	gaagaaaaaa	tgggtaaaag	acaacaatag	43680
acatttcaca	tttaacctca	ttcatgataa	ggtaagtga	aatgaaaact	acaggggata	43740
cctttttttt	tttttaatcc	attagattgg	caaacatccc	aaggtttgat	cataggctca	43800
gtgggtgaga	tttaagtatt	atcaggcatt	tttatacttt	gctgttagga	atgcaatgta	43860
gtacaaaacct	ttgtagaagt	tgcttttgaa	atgtctctca	gatgtacaaa	tgcatctaca	43920
tttttagattt	agcattcccg	ctttctgaga	cattattcaa	catgtatacg	tgtgcacata	43980
agatataata	ataacacgtt	tttccctcta	gtgtgttgct	tttaacctgt	agcttgaaaa	44040
aactctgctt	tcattgtttt	tttttgtttt	ctgtcactgg	ctcagccctg	ctttcaattg	44100
ttttatgaa	ttgatgggtg	ttctgggtctg	gttataatct	actttagttt	aagagtcaact	44160
ttaaattata	tgacatctga	tataagttgt	gttaggtaga	aaattctgta	acttggaata	44220
ctgtaagtac	tttgtggcca	catttcatta	gtattaaata	ttatctctat	atatagtagg	44280
ctattttaata	ttcatatttt	atgatgcaat	taagaaataa	ttttttctg	aagttggtag	44340
attgttgata	tgccatggcc	cagtgtttct	caaagcattc	tgggggatca	ctggtttgtca	44400
gaattagctg	cagtgtattgt	tgaacatgca	gggcctctgc	tccactccac	gttgctacca	44460
ggacgctctg	caggtgagag	ctgggaagct	gtagaagctg	cagtgtcaac	aaatgtctca	44520
ggaattcttg	tagtcacctt	catgaggtct	tatgttgagg	agaggcagcc	agtagtgctc	44580
cctgtccttc	ccgtttttatg	gtgtaagttt	catttttaagg	gaggtataaa	tcaaaagcca	44640
cctgggcatt	ctctcatggt	tactgtcttc	ttgtaatcat	ggaagatgtc	attgctggcag	44700
agacgaaaca	gtgtagtttg	attactattg	attttttttt	aattattttt	ctgaagtggc	44760
tggttgtaatg	taataaattg	tgtgcttaag	gacaaccttt	ggtattctat	ttgagtattg	44820
tgtatgatcc	tagtttaagt	ttttctacca	gtattttcat	attacaacat	atttactttc	44880
catttctatt	aatattttta	tattttaaagt	atggaggccg	ggcacagtg	ctcacgcgtg	44940
taatcccagc	attttgggat	gctgaggcgg	gtggatcaca	aggtcaggag	ttctagacca	45000
gcgtgaccaa	cacggtgaaa	tcccatctct	actaaaaata	caaaaattag	ccgggcacag	45060
tggtaggcac	ctgtaattcc	agctactcag	gaggctgagg	taggagaatc	acttgaatcc	45120

gggaggcagc	agttgcagtg	agctaagatc	gtgccactgg	actctagcct	ggctgacaga	45180
gcaagaatcc	gcctaaaaaa	aaagggatca	gggaagaggg	gattacagat	aacccaaaga	45240
agaaggaaaa	atcfccacaa	gttcacctgt	ccagcggtaa	ccccaatgtg	gatattttcc	45300
tttaacaatt	tggatatatt	cctttaaatc	ctctttttta	taatgtctat	atgttgga	45360
gagtatgtgc	ctttacgtat	tttttaaaga	tgagatttct	gtgtgtgtct	atatctcctg	45420
ttcttcatat	tttcttgtgt	gttataaaca	gctgtacatg	tcagtatata	tacttccgta	45480
actttttttt	aaaggctata	tagtgttcat	tgatgtgatt	taacagcagt	tatctccccg	45540
gttctacctt	gttggaatgt	gggtccctgt	gtttgccttc	agagcaaatg	gggcttggtt	45600
ttgacagcaag	tagacctgtg	acctgtacga	atagttggaa	gactttctct	attaccgaag	45660
cgtatcagta	tacttttagtg	cctactagaa	atttatgggt	agaaaaacaa	taatatctta	45720
gagtatTTTT	tcctagattc	cctaaggtgc	tatagggtga	tttttactca	tgtaacatga	45780
actatgcttc	aactaagata	gtttttgcaa	atgtggatat	ataagtactt	tattaaacct	45840
ataggaagta	tttataccac	ttatttctct	ccttcagtg	tagaacctcc	taaatggcat	45900
ttgacattga	actgctttcc	actttgtcgc	atgctcctct	cattgtccct	acctgggtcc	45960
tgaaccttag	ggacttggt	gttatagccc	caccatggct	acgctgggct	ttggctgtct	46020
gctgacttta	gtttcttcat	cttacaagga	gataataaca	gccccctgct	cgtagaatt	46080
gcagagatca	aatgaaataa	ttaacatact	caaaagcatg	ccgtaaacac	attctgagca	46140
catgtacgtt	ttaggaaaaa	caaaaggacc	catgcacatt	tcggagtgtc	tttgtctcag	46200
cagcactgcc	tcttcttcca	aagctgacgt	cttagtagag	gccctggcac	gtcctgagca	46260
ctgtactcca	cgaagcattc	tatttctgac	attcgaaatg	cagtctgttc	catcttccct	46320
acaatctgta	tgccagcact	tgaataaccg	ggatatctga	gtgttgacca	ggtgattact	46380
taattatgga	aatgttgagg	tggagatcta	gataattcag	tgaaggcagg	aaaattgggtg	46440
tcggaatctg	tctttttatg	tgtcagaaat	agaaataaga	taggggtgaga	agtaatttgt	46500
ggctaaaaaca	ctataatagc	taacacatag	tgcatactgt	gtgccaaagca	ctcctgtagg	46560
tgcttgaaat	cttctattat	tattatccct	actttataga	cttgaccctt	taggcacaga	46620
gaggcgagca	gttgtccaag	gttaccctag	agggtggagat	ccaggctacc	tgactccacc	46680
atgtgtgtct	ttccctaggg	cacagtgtgt	ctgctaaaaa	tactttttta	gcagttcttt	46740
gattattcag	atgatatgtc	tgtaggaaaa	ttaagacaaa	aataatgaaa	aatataaatc	46800
tttatttttag	tgttttgcac	atgtattatt	aaagccagtt	tactcctgga	agtgtgtaag	46860
aatacaggg	atttttgatc	acctaaatgc	tgcatgttac	taagagctcg	acactgaagt	46920
caagaagagc	agttgcagag	agtacttagc	aaaaacggga	agtgtgtggg	gttgaaggag	46980
caaagacaag	tcttctctcg	acggtggagt	tgagaattca	tcatttctca	gaacacgtct	47040
ttgaacgcat	tttcaatttg	aggccaaagg	tctcagcctc	ccactcggca	tacctcccta	47100
ccttagtcag	ctcttaaatc	ttaggaatat	ttctttgttc	ttcaagggaac	ttaaatatgt	47160
taacattctt	acctgtccac	agggagcccc	ctacaaagaa	gggagtttct	agtcctcggt	47220
cttctcttga	ataaataata	gcctcatacc	ttgtgcaatc	gaggctgata	aagactgtct	47280
ccttttttca	aataagcaag	tcttagaaac	tacagttgtt	tacagggctc	atggctattc	47340
cacagtaata	atttttggtc	ttttaccaat	tatataatat	gttaaaatat	ggcaagtatc	47400
aggaaagcaa	ggagtggcaa	tgattagaaa	ccaatggcca	agttagagag	gaggggcaat	47460
tgctccccc	agtttggtgt	ggctgtgtag	cagtcagtga	cgagaagctg	gtgtgcaggc	47520
gacaagcaaa	gttgaggatt	atcaggcgcc	tgtgagtgcc	cagctgtgtg	ccaggtcagg	47580
agggtgccatc	gtgagccaga	ccagcttctc	ctcggccctc	gtggagctcg	cagctcgggtg	47640
gggaggcagc	agtcaccatg	gtgacagggt	acacactagg	atggggctgg	tggtggtagg	47700
catttgcggt	tcccttcaga	gaggtgagta	tggacttaga	ggaggctcca	gcttctatt	47760
cctgggctgt	ctatagcact	aaaagtgtgc	acatgaaaaa	taacatttgg	tactattgat	47820
ttactttaat	gacttatgta	attgtagtgt	acttagaaat	tataacatgc	tcttctactt	47880
cagcttgaaa	cccccaacca	ccagtttata	atcctttttt	tttaactttt	gtttattttt	47940
cctaaggaat	ctgtactttt	tcttcatttt	acaacttttt	ttgtcctgtt	accttatttt	48000
cattttttact	ttatatgacc	atgagttcta	aaatagtaaa	aaaaaagaat	tattttttgt	48060
ctttgttaga	atttctctgc	aaagaatgtc	caaaaattca	tattcacatt	gatcgtatcg	48120
acaaaaaaga	tgtcccagaa	gaacaagaac	atatgagaag	atggctgcat	gaacgtttcg	48180
aaatcaaaga	taagttagta	acaacagttc	cagcacttcc	ggaacttcgg	ttcaactaga	48240
tttcagttata	gtcaacaatt	tgaaccaaat	gtaaaatggt	atattgtctc	aagaatacat	48300
tttataaatt	caaatcaaat	tttatgcatg	tctgatcgtg	ttttaaactt	tacttgtaca	48360
aatcagttcta	aaagaacttg	ttacagtggg	cccatctact	tgcatgtata	gtatttcttg	48420
gacaatacta	cgtgataaca	tagcaaatca	aattaaaaac	aacaacaaac	acacaaaaaa	48480
actttccagt	gtcagatgcc	cggacctacc	tgtcaggtca	cataaagtgg	tggtactgtg	48540
tgagggtctg	ctgttggggc	agtgtgcgca	gaaaagcaag	ggaggggtag	aggactatgc	48600
ggcagtgtag	gtggacatga	tgctgttata	tttgttggaa	atagaagggg	gcagttgaca	48660
gcgttatatc	caaagtgtct	tctgtgttta	attatatcca	gaaatttttag	ccaattgttt	48720
tattctctaa	atatgtactt	tctgtctcaag	aaactatcat	tggtcttctt	ttccttggtt	48780

tacagtagac tgtttttaat taaccctcct gggttaactt taccaggtga aaatgattaa 48840
aagtgtataa ggttaacaat gaaactttta gcttctattt ttcattgact cttaactgta 48900
catgatgtaa tgtatttcagc gagccattca ggaccacttt ggcccatgga agaaatttaa 48960
aagtaagatc tacatgtatt gacatgaaaa tatgtttctca gaaaaaagac taatgtattt 49020
aatgtccctac ttattttata agtattttaga atacctctgg acatttttaa acaatgatta 49080
ttgctagggg gtgtgattta taaagcaata gaagcgcttt ccctttctgt ttgtgtttta 49140
gattattata tcgggtatgt tctgctatca taactttaca aatcttatgt aatatgggaa 49200
aatgagttaa ctatgctgtt ttcttctttt tacctgcctt tctaattctg tgggaataaa 49260
ggcggttttg agacagccca ggtgtagtga gcagtcata tccatggatt ccacattcat 49320
ggattccacc aagcacagac caaaaatact cagaaaaaaa gggggctggc tgtgtgggt 49380
catgcatgta atcccagcac tttgggaggg taaggcaggg aaattgcttg agcccgaga 49440
ttcaagacag cctgggcaac atggcaaaac cctgtctcta cagaaaaata aaaaattagc 49500
caggcgtgca cctgtagtcc cagctactca ggaggccgag gtgcgaggat cacctgagcc 49560
tgggaagggt agactgcagt gagctatcat tgtgccaaact ccagcctggg aacagagtgc 49620
cttttttcaa aaaaaaaaaa aaaaaaggat ttgggaggat atgcatatgt tatattcaa 49680
tacatgccat tttattcata tatcagggac ttgagcatcc tttgatcttg gtctctgccg 49740
gggtatcctgg gaccagcccc ctgtcgatac agagggaccg ctgtctaaaga accgctgggtc 49800
ctatctttga ctctctggcg aataggagct ccatgtaaaa aggaggagaa gctgcagcgg 49860
gttattagcc atttctgagt caggtcactg taaaacttta tcaaaaagttt aaaagacaaa 49920
aagcatcctc ataaaaatgcc ttaaaaccac ctgtgaaat attacatata caattcatgt 49980
atactaatca tagagcatat taaagatatt ttagaagact agaaacttct attaaacca 50040
gtttctggat gtttccgtat tcatccttat ttccaggga cctgcataac tttccagcg 50100
tgtaaatagct acctgattga tattttttga attgaaatcc tgaagtgtact aaaatctaaa 50160
ctttttccat tctggccata ggatgcttat agaattttat ggtcaccag atccagaaaag 50220
aagaaaaaga tttcctggga aaagtgttaa ttccaaatta agtatcaaga agactttacc 50280
atcaatgttg atcttaagtgc gtttgactgc aggcattgctt atgaccgatg ctggaaggaa 50340
gctgtatgtg aacacctgga tatatggaac cctacttggc tgcctgtggg ttactattaa 50400
agcatagaca agtagctgtc tccagacagt gggatgtgct acattgtcta tttttggcg 50460
ctgcacatga catcaaatg tttcctgaat ttattaagga gtgtaataa agccttgttg 50520
attgaagatt ggataataga atttgtgacg aaagctgata tgcaatggct tgggcaaac 50580
atacctggtt gtacaacttt agcatcgggg ctgctggaag ggtaaaagct aaatggaagt 50640
tctcctgctc tgtccatttc ctatgaacta atgacaactt gagaaggctg ggaggattgt 50700
gtattttgca agtcagatgg ctgcattttt gagcattaat ttgcagcgta ttccactttt 50760
tctgttattt tcaatttatt acaacttgac agctccaagc tcttattact aaagtattta 50820
gtatcttgca gctagttaat atttcatctt ttgcttattt ctacaagtca gtgaaataa 50880
ttgtatttag gaagtgtcag gatgttcaaa ggaaagggtta aaaagtgttc atggggaaaa 50940
agctctgttt agcacatgat tttattgtat tgcgttatta gctgatttta ttcattttat 51000
atttgcaaaa taaatttcta atattttatt aaattgctta atttgcacac cctgtacaca 51060
cagaaaaatgg tataaaatat gagaacgaag tttaaaattg tgactctgat tcattatagc 51120
agaactttta atttccagc tttttgaaga ttttaagctac gctattagta cttccctttg 51180
tctgtgccat aagtgttga aaacgttaag gttttctgtt ttgttttgtt tttttaatat 51240
caaaagagtc ggtgtgaacc ttggttgagc cccaagttca caagattttt aaggtgatga 51300
gagcctgcag acattctgcc tagatttact agcgtgtgct ttttgctgc tttctcttga 51360
tttcacagaa tattcattca gaagtcgctt ttctgtagtg tgggtggattc ccactgggct 51420
ctggctcctc ccttggatcc cgtcagtggt gctgctcagc ggcttgcagc tagacttgct 51480
aggaagaaat gcagagccag cctgtgctgc ccactttcag agttgaaact ttttaagcctt 51540
tgtgagtggt cttcaccagc tactgcagag gcattttgca tttgtctgtg tcaagaagtt 51600
caccttctca agccagtgaa atacagactt aattcgtcat gactgaacga atttgtttat 51660
ttcccattag gtttagtgga gctacacatt aatatgtatc gccttagagc aagagctgtg 51720
ttccaggaac cagatcacga tttttagcca tggacaata tatcccatg gagaagacct 51780
ttcagtggtg actgttctat ttttgtgta taatttaaac ttcgatttcc tcatagtctt 51840
ttaagttgac atttctgctt actgctactg gatttttgcg gcagaaatat atcagtggcc 51900
cacattaaac ataccagtgt gatcatgata agcaaaatga aagaaataat gattaaggga 51960
aaattaaagt actgtgttac actgcttctc ccatgccaga gaataaaact tttcaagcat 52020
catctttgaa gagtcgtgtg gtgtgaattg gtttgtgtac attagaatgt atgcacacat 52080
ccatggacac tcaggatata gttggcctaa taatcggggc atgggtaaaa cttatgaaaa 52140
tttctctcat ctgaattgta atttctctt accgtgaaag taaaatttag atcaattcca 52200
tgtctttgtt aagtacaggg atttaataata ttttgaatat aatgggtatg ttctaaattt 52260
gaactttgag aggcataact gttggaatta tgtggattct aactcatttt aacaaggtag 52320
cctgacctgc ataagatcac ttgaatgtta ggtttcatag aactatacta atcttctcac 52380
aaaaggtcta taaaatacag tctgtgaaaa aaattttgta tcaaaatgtt tggaaaaata 52440

gaagcttctc	cttaacctgt	attgatactg	acttgaatta	ttttctaaaa	ttaagagccg	52500
tatacctacc	tgtaagtctt	ttcacatatc	atttaaactt	ttgtttgtat	tattactgat	52560
ttacagctta	gttattaatt	ttcttttata	agaatgccgt	cgatgtgcat	gctttttatgt	52620
ttttcagaaa	aggggtgtgt	tggtatgaaag	taaaaaaaaa	aataaaatct	ttcactgtct	52680
ctaattggctg	tgctgtttaa	cattttttga	ccctaaaatt	caccaacagt	ctcccagttac	52740
ataaaatagg	cttaatgact	ggccctgcat	tcttcacaat	atTTTTccct	aagctttgag	52800
caaagtttta	aaaaaataca	ctaaaataat	caaaactgtt	aagcagtata	ttagtttggg	52860
tataataaatt	catctgcaat	ttataagatg	catggccgat	gttaatttgc	ttggcaattc	52920
tgtaatcatt	aagtgatctc	agtgaacat	gtcaaatgcc	ttaaattaac	taagttgggtg	52980
aataaaagtg	ccgatctggc	taactcttac	accatacata	ctgatagttt	ttcatatgtt	53040
tcatttccat	gtgattttta	aaatttagag	tggaacaat	tttgcttaat	atgggttaca	53100
taagctttat	tttttccctt	gttcataatt	atattctttg	aataggtctg	tgtcaatcaa	53160
gtgatctaac	tagactgatc	atagatagaa	ggaaataagg	ccaagttaaa	gaccagcttg	53220
ggcaacatat	cgagaacctg	tctacaaaaa	aattaaaaaa	aattagccag	gcattggtggc	53280
gtacactgag	tagttttgtcc	cagctactcg	ggaggggtgag	gtgggaggat	cgcttcagcc	53340
caggaggttg	agattgcagt	gagccatgga	cataccactg	cactacagcc	taggtaacag	53400
cacgagaccc	caactcttag	aaaatgaaaa	ggaaatatag	aaatataaaa	tttgcttatt	53460
atagacacac	agtaactccc	agatatgtac	cacaaaaaat	gtgaaaagag	agagaaatgt	53520
ctaccaaaagc	agtattttgt	gtgtataaatt	gcaagcgcag	agtaaaataa	ttttaacctt	53580
aatttgTTTT	tagtagtggt	tagattgaaag	attgagtga	atattttctt	ggcagatatt	53640
ccgtatctgg	tggaagacta	caatgcaatg	tcgttgtagt	tttgcatggc	ttgctttata	53700
aacaagattt	tttctccctc	cttttgggcc	agttttcatt	acgagtaact	cacacttttt	53760
gattaaagaa	cttgaaattt	cgttatcact	tagtataaatt	gacattatat	agagactatg	53820
taacatgcaa	tcattagaat	caaaaattag	actttgggtca	aaatatttac	aacattcaca	53880
tacttgctcaa	atattcatgt	aattaaactga	attttaaacc	ttcaactatt	atgaagtgc	53940
cgtctgtaca	atcgctaatt	tactcagttt	agagttagcta	caactcttcg	atactatcat	54000
caatatTTTga	catcttttcc	aatttgtgta	tgaaaagtaa	atctattcct	gtagcaactg	54060
gggagtcata	tatgagggtca	aagacatata	ccttgttatt	ataatatgta	tactataata	54120
atgctgggtt	atcctgagca	ggggaagagg	ttatttttag	gaaaaccact	tcaaatagaa	54180
agctgaagta	cttctaatat	actgagggaa	gtataaatatg	tggaacaaac	tctcaacaaa	54240
atgtttattg	atgttgatga	aacagatcag	tttttccatc	cggattatta	ttgggtcatg	54300
atTTTtatag	tgaatatgta	agatatgttc	tgcaatttta	taaattgtca	gtcttttttt	54360
taaaaaaggt	gctattgaaa	ttctgtgtct	ccagcaggca	agaatacttg	actaactctt	54420
tttgtctctt	taTgggtattt	tcagaataaa	gtctgacttg	tgTTTTtgag	attattgggtg	54480
cctcattaat	tcagcaataa	aggaaaaatat	gcattctcaa	aattgggtgat	aaaaagttat	54540
ttcttgTata	tgtgataaag	tttcatgtgt	gtgtatatat	gttgatttgc	caaatacggc	54600
ttttaaatac	taPgTcatat	tttaaagggt	cagttttag	tgatagtaaa	caagcagtg	54660
actaagcctc	ttgcgggcat	catctcatct	cactgtcatc	acaaacccca	tgccacagcg	54720
tagcttgacc	actaaaagta	atgcactctg	aagcactactg	ccagggtttg	gatagtttgt	54780
accaacagtg	accttatcaa	ggtaaaatccc	agactctaaa	agagttgggtg	ctgtgtcact	54840
acatgcataa	ctttaaataa	atttcctgcc	gggcgcgggtg	gctcacgcct	gtaatcccag	54900
cagtttgagg	ggccgaggca	agtggatcac	ttgaggtcag	gagtttgaga	ccagcctggc	54960
caacgtgggtg	aaacccgtgc	tctactaaaa	atacaaaaat	tagccaggcg	tgtgggtggca	55020
ggcacctgta	atcccagcta	cttgggagga	tgaggcagga	gaatcatttg	aatcctgcag	55080
gcggagggtg	cagtggagcca	agatggcgctc	attgcactcc	agcctggggcg	acaagagcga	55140
gactccgtat	taaaaaaaaa	aaaaaaaaaa	aaaaaaaaatt	cctctcctgt	ttgagctttc	55200
ccttacctgt	aaagagggga	gaatatgtat	ttacttcaa	gagttcagg	aaatgactct	55260
cactagtttg	agattctagg	tataaaaaata	cattcttata	taattttaac	accaatgtga	55320
gagattatta	ttcttgctaa	accaattcag	ttttatttgc	tgtctaaaat	gtgtgaataa	55380
gtAattgtcc	attattttct	gaagtgtttt	ggaactcaac	acatgattgt	gaggaggatt	55440
tgttgctaaa	catctttctg	gttattcaag	ctcgtgtata	ctgtgctctg	ttgagacatg	55500
cagagttact	ttctgtctgg	gtcacagggtc	agttcttgat	agttttcgga	caattaacca	55560
gttttcattt	gccccatgac	acctttatc	tttttctca	actgcacca	tcttttataa	55620
ggtctttcag	tttattgcag	agaagatgg	ggagaaaagc	cggaattccc	acccaccgct	55680
gccatccccca	tgttttatca	ttggctagag	tggaataatg	cagtaactac	tgtgagagat	55740
catttgTTTT	tataatggaa	acaaagatga	ggaaagaacc	tggcttagat	cagagaactg	55800
atgtatttag	attctttttt	ttttttttt	taagacggag	tggtgctctg	ttgccagac	55860
tgaggtacag	tggtcctaatc	tcggctcact	gcaacctcca	tttccctgg	tcaagcaatt	55920
atcctgcctc	agcctcccaa	gtatttggga	ttacaggcgt	gttccaccac	acctggctaa	55980
ttttttgtat	tttttagtaga	gacgggggtt	cgccatgttg	gccaggctgg	tctcgaaatc	56040
ctgacctcag	atgatccacc	cgccttggcc	tcccaaagt	ctgggattac	aggcgcgagc	56100

```

caccgcgcct ggcccaatgt atttggattc ttaaagaaca ctttcaaatt aaatatcagt 56160
tgaagagAAC tagaactaaa gaatttctgt gtcaaaactgt ttagcaaatg taagtagaag 56220
ctgggagatg tgtctctgga tgaatgaata catcagtaaa ataccatacg tatgttatga 56280
tggtattgtt tctttgcctt gggttgattg gttttactgt gaaataattt tcaatataga 56340
attgtgatcg ttggaatttg gtcactctact agaaaatgag aaagaagtta atagctatct 56400
tccttaaaga tttctgaggt tgggattaag gtagtggtcc caaggtgttc taaaacggca 56460
gcgagagctg tgcactcact tcacaaattt gaattcctgc tctgtgttag gcgctg 56516
<210> 2
<211> 23
<212> DNA
<213> Homo sapiens
<400> 2
ggtcgtccag cgcttggttag aag 23
<210> 3
<211> 5227
<212> DNA
<213> Homo sapiens
<220>
<221> polyA_signal
<222> 5180..5186
<223> AATAAA
<400> 3
ctgctgtccc tgggtgtcca cagctactcc atg cgc tac ctg ctg ccc agc gtc 54
Met Arg Tyr Leu Leu Pro Ser Val
1 5
gtg ctc ctg ggc acg gcg ccc acc tac gtg ttg gcc tgg ggg gtc tgg 102
Val Leu Leu Gly Thr Ala Pro Thr Tyr Val Leu Ala Trp Gly Val Trp
10 15 20
cgg ctg ctc tcc gcc ttc ctg ccc gcc cgc ttc tac caa gcg ctg gac 150
Arg Leu Leu Ser Ala Phe Leu Pro Ala Arg Phe Tyr Gln Ala Leu Asp
25 30 35 40
gac cgg ctc tac tgc gtc tac cag agc atg gtg ctc ttc ttc ttc gag 198
Asp Arg Leu Tyr Cys Val Tyr Gln Ser Met Val Leu Phe Phe Phe Glu
45 50 55
aat tac acc ggg gtc cag ata ttg cta tat gga gat ttg cca aaa aat 246
Asn Tyr Thr Gly Val Gln Ile Leu Leu Tyr Gly Asp Leu Pro Lys Asn
60 65 70
aaa gaa aat ata ata tat tta gca aat cat caa agc aca gtt gac tgg 294
Lys Glu Asn Ile Ile Tyr Leu Ala Asn His Gln Ser Thr Val Asp Trp
75 80 85
att gtt gct gac atc ttg gcc atc agg cag aat gcg cta gga cat gtg 342
Ile Val Ala Asp Ile Leu Ala Ile Arg Gln Asn Ala Leu Gly His Val
90 95 100
cgc tac gtg ctg aaa gaa ggg tta aaa tgg ctg cca ttg tat ggg tgt 390
Arg Tyr Val Leu Lys Glu Gly Leu Lys Trp Leu Pro Leu Tyr Gly Cys
105 110 115 120
tac ttt gct cag cat gga gga atc tat gta aag cgc agt gcc aaa ttt 438
Tyr Phe Ala Gln His Gly Gly Ile Tyr Val Lys Arg Ser Ala Lys Phe
125 130 135
aac gag aaa gag atg cga aac aag ttg cag agc tac gtg gac gca gga 486
Asn Glu Lys Glu Met Arg Asn Lys Leu Gln Ser Tyr Val Asp Ala Gly
140 145 150
act cca atg tat ctt gtg att ttt cca gaa ggt aca agg tat aat cca 534
Thr Pro Met Tyr Leu Val Ile Phe Pro Glu Gly Thr Arg Tyr Asn Pro
155 160 165
gag caa aca aaa gtc ctt tca gct agt cag gca ttt gct gcc caa cgt 582
Glu Gln Thr Lys Val Leu Ser Ala Ser Gln Ala Phe Ala Ala Gln Arg
170 175 180
ggc ctt gca gta tta aaa cat gtg cta aca cca cga ata aag gca act 630
Gly Leu Ala Val Leu Lys His Val Leu Thr Pro Arg Ile Lys Ala Thr
185 190 195 200

```

cac gtt gct ttt gat tgc atg aag aat tat tta gat gca att tat gat	678
His Val Ala Phe Asp Cys Met Lys Asn Tyr Leu Asp Ala Ile Tyr Asp	
205 210 215	
ggt acg gtg gtt tat gaa ggg aaa gac gat gga ggg cag cga aga gag	726
Val Thr Val Val Tyr Glu Gly Lys Asp Gly Gly Gln Arg Arg Glu	
220 225 230	
tca ccg acc atg acg gaa ttt ctc tgc aaa gaa tgt cca aaa att cat	774
Ser Pro Thr Met Thr Glu Phe Leu Cys Lys Glu Cys Pro Lys Ile His	
235 240 245	
att cac att gat cgt atc gac aaa aaa gat gtc cca gaa gaa caa gaa	822
Ile His Ile Asp Arg Ile Asp Lys Lys Asp Val Pro Glu Glu Gln Glu	
250 255 260	
cat atg aga aga tgg ctg cat gaa cgt ttc gaa atc aaa gat aag atg	870
His Met Arg Arg Trp Leu His Glu Arg Phe Glu Ile Lys Asp Lys Met	
265 270 275 280	
ctt ata gaa ttt tat gag tca cca gat cca gaa aga aga aaa aga ttt	918
Leu Ile Glu Phe Tyr Glu Ser Pro Asp Pro Glu Arg Arg Lys Arg Phe	
285 290 295	
cct ggg aaa agt gtt aat tcc aaa tta agt atc aag aag act tta cca	966
Pro Gly Lys Ser Val Asn Ser Lys Leu Ser Ile Lys Lys Thr Leu Pro	
300 305 310	
tca atg ttg atc tta agt ggt ttg act gca ggc atg ctt atg acc gat	1014
Ser Met Leu Ile Leu Ser Gly Leu Thr Ala Gly Met Leu Met Thr Asp	
315 320 325	
gct gga agg aag ctg tat gtg aac acc tgg ata tat gga acc cta ctt	1062
Ala Gly Arg Lys Leu Tyr Val Asn Thr Trp Ile Tyr Gly Thr Leu Leu	
330 335 340	
ggc tgc ctg tgg gtt act att aaa gca tag acaagtagct gtctccagac	1112
Gly Cys Leu Trp Val Thr Ile Lys Ala *	
345 350	
agtgggatgt gctacattgt ctatttttgg cggtgcaca tgacatcaaa ttgtttcctg	1172
aatttattaa ggagtgtaaa taaagccttg ttgattgaag attggataat agaatttgtg	1232
acgaaagctg atatgcaatg gtcttgggca aacatacctg gttgtacaac tttagcatcg	1292
gggctgctgg aagggtaaaa gctaaatgga gttttcctcg ctctgtccat ttcctatgaa	1352
ctaatacaca cttgagaagg ctgggaggat tgtgtatttt gcaagtcaga tggctgcatt	1412
tttgagcatt aatttgcagc gtatttcact ttttctgtta ttttcaattt attacaactt	1472
gacagctcca agctcttatt actaaagtat ttagtatctt gcagctagtt aatatttcatt	1532
cttttgcctta tttctacaag tcagtgaat aaattgtatt taggaagtgt caggatgttc	1592
aaaggaaaagg gtaaaaagtg ttcattggga aaaagctctg tttagcacat gattttattg	1652
tattgcgtta ttagctgatt ttactcattt tatatttgca aaataaattt ctaatattta	1712
ttgaaattgc ttaatttgca caccctgtac acacagaaaa tggataaaaa tatgagaacg	1772
aagtttaaaa ttgtgactct gattcattat agcagaactt taaatttccc agctttttga	1832
agattttaagc tacgctatta gtactccct ttgtctgtgc cataagtgct tgaaaacgtt	1892
aagggtttct gttttgtttt gtttttttaa tatcaaaaaga gtcggtgtga acctgggtg	1952
gaccccaagt tcacaagatt tttaagggtga tgagagcctg cagacattct gcctagattt	2012
actagcgtgt gccttttgcc tgcttctctt tgatttcaca gaatttcat tcagaagtcg	2072
cgttttctgta gtgtggtgga ttcccactgg gctctgtgct ttccttgga tcccgtcagt	2132
ggtgctgctc agcggcttgc acgtagactt gctaggaaga aatgcagagc cagcctgtgc	2192
tgcccacttt cagagttgaa ctcttttaag cccttgtgag tgggcttcac cagctactgc	2252
agaggcattt tgcatttgtc tgtgtcaaga agttcacctt ctcaagccag tgaaatacag	2312
acttaattcg tcatgactga acgaatttgt ttatttccca ttaggtttag tggagctaca	2372
cattaatatg tatgcctta gagcaagagc tgtgttccag gaaccagatc acgattttta	2432
gccatggaac aatatatccc atgggagaag acctttcagt gtgaactgtt ctatttttgt	2492
gttataattt aaacttcgat ttctcatag tcctttaagt tgacatttct gcttactgct	2552
actggatttt tgctgcagaa atatatcagt ggcccacatt aaacatacca gttggatcat	2612
gataagcaaa atgaaagaaa taatgattaa gggaaaatta agtgactgtg ttacactgct	2672
tctcccatgc cagagaataa actctttcaa gcatcatctt tgaagagtcg tgtggtgtga	2732
attggtttgt gtacattaga atgtatgcac acatccatgg acactcagga tatagttggc	2792
ctaataatcg gggcatgggt aaaacttatg aaaatttcct catgctgaat tgtaattttc	2852
tcttacctgt aaagtaaaat ttagatcaat tccatgtctt tgttaagtac agggatttaa	2912
tatattttga atataatggg tatgttctaa atttgaactt tgagaggcaa tactgttgga	2972

```

attatgtgga ttctaactca ttttaacaag gtagcctgac ctgcataaga tcacttgaat 3032
gttaggtttc atagaactat actaatcttc tcacaaaagg tctataaaat acagtcggtg 3092
aaaaaaatgt tgtatcaaaa tggttggaag attagaagct tctccttaac ctgtattgat 3152
actgacttga attattttct aaaattaaga gccgtatacc tacctgtaag tcttttcaca 3212
tatcatttaa acttttgttt gtattattac tgatttacag cttagttatt aattttctct 3272
tataagaatg ccgctgatgt gcatgctttt atgtttttca gaaaagggtg tggttgatg 3332
aaagtaaaaa aaaaaataaa atctttcact gtctctaag gctgtgctgt ttaacatttt 3392
ttgaccctaa aattcaccaa cagtctccca gtacataaaa taggcttaat gactggccct 3452
gcattcttca caatattttt ccctaagctt tgagcaaaagt tttaaaaaaa tacactaaaa 3512
taatcaaaac tgtaagcag tatattagtt tggttatata aattcatctg caatttataa 3572
gatgcatggc cgatgttaat ttgcttgga attctgtaat cattaagtga tctcagtga 3632
acatgtcaaa tgccttaaat taactaagtt ggtgaataaa agtgccgac tggtcaactc 3692
ttacaccata catactgata gtttttcata tgtttcattt ccatgtgatt tttaaaattt 3752
agagtggaac caattttgct taatatgggt tacataagct ttattttttc cttgttcat 3812
aatttatatt tttgaatagg tctgtgtcaa tcaagtgatc taactagact gatcatagat 3872
agaaggaaat aaggccaaagt tcaagaccag cctgggcaac atatcgagaa cctgtctaca 3932
aaaaaattaa aaaaaattag ccaggcatgg tggcgtacac tgagttagtt gtcccagcta 3992
ctcgggaggg tgaggtggga ggatcgcttc agccaggag gttgagattg cagtgaacca 4052
tggaacatac actgcactac agcctaggta acagcacgag accccaactc ttagaaaatg 4112
aaaaggaaat atagaaatat aaaatttgct tattatagac acacagtaac tcccagatat 4172
gtaccacaaa aaatgtgaaa agagagagaa atgtctacca aagcagtatt ttgtgtgat 4232
aattgcaagc gcatagtaaa ataattttta ccttaatttg tttttagtag tgtttagatt 4292
gaagattgag tgaaatattt tcttggcaga tattccgtat ctggtggaaa gctacaatgc 4352
aatgtcgttg tagttttgca tggcttgctt tataaacaag attttttctc cctccttttg 4412
ggccagtttt cattacgagt aactcacact ttttgattaa agaacttgaa attacgttat 4472
cacttagtat aattgacatt atatagagac tatgtaacat gcaatcatta gaatcaaaat 4532
tagtactttg gtcaaaatat ttacaacatt cacatacttg tcaaataatc atgtaattaa 4592
ctgaatttaa aaccttcaac tattatgaag tgctcgtctg tacaatcgct aatttactca 4652
gtttagagta gctacaactc ttcgatacta tcatcaatat tggacatctt ttccaatttg 4712
tctatgaaaa gtaaatctat tctgttagca actggggagt catatatgag gtcaaagaca 4772
tataccttgt tattataata tgtatactat aataatagct ggttatcctg agcaggggaa 4832
aagggttatt ttaggaaaaac cacttcaaat agaaagctga agtacttcta atatactgag 4892
ggaagtataa tatgtggaac aaactctcaa caaaatgttt attgatgttg atgaaacaga 4952
tcagtttttc catccggatt attattgggt catgatttta tatgtgaata tgtaagatat 5012
gttctgcaat tttataaatg ttcattgtct tttttaaaaa aggtgctatc gaaattctgt 5072
gtctccagca ggcaagaata cttgactaac tctttttgtc tctttatggt attttcagaa 5132
taaagtctga cttgtgtttt tgagattatt ggtgcctcat taattcagca ataaaggaaa 5192
atatgcattt caaaaaanaa aaaaaaaaaa aaaaaa 5227

```

<210> 4

<211> 353

<212> PRT

<213> Homo sapiens

<220>

<221> HELIX

<222> 1..33

<223> Rao and Argos identification method, potential helix

<221> HELIX

<222> 4..20

<223> Klein, Kanehisa and DeLisi identification method, potential helix

<221> HELIX

<222> 4..24

<223> Eisenberg, Schwarz, Komarony, Wall identification method,

potential helix

<221> MYRISTATE

<222> 12..16

<223> Prosite match

<221> HELIX

<222> 50..70

<223> Eisenberg, Schwarz, Komarony, Wall identification method,

potential helix

<221> CARBOHYD

<222> 57..59
<223> Prosite match
<221> HELIX
<222> 76..96
<223> Eisenberg, Schwarz, Komarony, Wall identification method,
potential helix
<221> PHOSPHORYLATION
<222> 78
<223> potential Tyrosine kinase site, Prosite match
<221> PHOSPHORYLATION
<222> 84
<223> potential caseine kinase II site, Prosite match
<221> SITE
<222> 94..115
<223> potential Leucine zipper site, Prosite match
<221> MYRISTATE
<222> 119..123
<223> potential site, Prosite match
<221> PHOSPHORYLATION
<222> 133
<223> potential protein kinase C, Prosite match
<221> PHOSPHORYLATION
<222> 147
<223> potential caseine kinase II site, Prosite match
<221> PHOSPHORYLATION
<222> 194
<223> potential protein kinase C, Prosite match
<221> PHOSPHORYLATION
<222> 215
<223> potential Tyrosine kinase site, Prosite match
<221> SULFATATION
<222> 221
<223> Prosite match
<221> PHOSPHORYLATION
<222> 233
<223> potential cAMP and cGMP dependant protein kinase site, Prosite
match
<221> PHOSPHORYLATION
<222> 235
<223> potential caseine kinase II site, Prosite match
<221> PHOSPHORYLATION
<222> 306
<223> potential protein kinase C, Prosite match
<221> HELIX
<222> 310..330
<223> Eisenberg, Schwarz, Komarony, Wall identification method,
potential helix
<221> MYRISTATE
<222> 319..323
<223> Prosite match
<221> MYRISTATE
<222> 323..327
<223> Prosite match
<221> AMIDATION
<222> 329
<223> Prosite match
<221> HELIX
<222> 333..353
<223> Eisenberg, Schwarz, Komarony, Wall identification method,
potential helix
<221> MYRISTATE

<222> 341..345

<223> Prosite match

<221> PHOSPHORYLATION

<222> 350

<223> potential protein kinase C, Prosite match

<400> 4

```

Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
1      5      10      15
Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
20     25     30
Ala Arg Phe Tyr Gln Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln
35     40     45
Ser Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln Ile Leu
50     55     60
Leu Tyr Gly Asp Leu Pro Lys Asn Lys Glu Asn Ile Ile Tyr Leu Ala
65     70     75     80
Asn His Gln Ser Thr Val Asp Trp Ile Val Ala Asp Ile Leu Ala Ile
85     90     95
Arg Gln Asn Ala Leu Gly His Val Arg Tyr Val Leu Lys Glu Gly Leu
100    105    110
Lys Trp Leu Pro Leu Tyr Gly Cys Tyr Phe Ala Gln His Gly Gly Ile
115    120    125
Tyr Val Lys Arg Ser Ala Lys Phe Asn Glu Lys Glu Met Arg Asn Lys
130    135    140
Leu Gln Ser Tyr Val Asp Ala Gly Thr Pro Met Tyr Leu Val Ile Phe
145    150    155    160
Pro Glu Gly Thr Arg Tyr Asn Pro Glu Gln Thr Lys Val Leu Ser Ala
165    170    175
Ser Gln Ala Phe Ala Ala Gln Arg Gly Leu Ala Val Leu Lys His Val
180    185    190
Leu Thr Pro Arg Ile Lys Ala Thr His Val Ala Phe Asp Cys Met Lys
195    200    205
Asn Tyr Leu Asp Ala Ile Tyr Asp Val Thr Val Val Tyr Glu Gly Lys
210    215    220
Asp Asp Gly Gly Gln Arg Glu Ser Pro Thr Met Thr Glu Phe Leu
225    230    235    240
Cys Lys Glu Cys Pro Lys Ile His Ile His Ile Asp Arg Ile Asp Lys
245    250    255
Lys Asp Val Pro Glu Glu Gln Glu His Met Arg Arg Trp Leu His Glu
260    265    270
Arg Phe Glu Ile Lys Asp Lys Met Leu Ile Glu Phe Tyr Glu Ser Pro
275    280    285
Asp Pro Glu Arg Arg Lys Arg Phe Pro Gly Lys Ser Val Asn Ser Lys
290    295    300
Leu Ser Ile Lys Lys Thr Leu Pro Ser Met Leu Ile Leu Ser Gly Leu
305    310    315    320
Thr Ala Gly Met Leu Met Thr Asp Ala Gly Arg Lys Leu Tyr Val Asn
325    330    335
Thr Trp Ile Tyr Gly Thr Leu Leu Gly Cys Leu Trp Val Thr Ile Lys
340    345    350

```

Ala

<210> 5

<211> 364

<212> PRT

<213> Homo sapiens

<400> 5

```

Met Leu Leu Ser Leu Val Leu His Thr Tyr Ser Met Arg Tyr Leu Leu
1      5      10      15
Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr Tyr Val Leu Ala Trp
20     25     30
Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro Ala Arg Phe Tyr Gln
340    345    350

```

35 40 45
 Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln Ser Met Val Leu Phe
 50 55 60
 Phe Phe Glu Asn Tyr Thr Gly Val Gln Ile Leu Leu Tyr Gly Asp Leu
 65 70 75 80
 Pro Lys Asn Lys Glu Asn Ile Ile Tyr Leu Ala Asn His Gln Ser Thr
 85 90 95
 Val Asp Trp Ile Val Ala Asp Ile Leu Ala Ile Arg Gln Asn Ala Leu
 100 105 110
 Gly His Val Arg Tyr Val Leu Lys Glu Gly Leu Lys Trp Leu Pro Leu
 115 120 125
 Tyr Gly Cys Tyr Phe Ala Gln His Gly Gly Ile Tyr Val Lys Arg Ser
 130 135 140
 Ala Lys Phe Asn Glu Lys Glu Met Arg Asn Lys Leu Gln Ser Tyr Val
 145 150 155 160
 Asp Ala Gly Thr Pro Met Tyr Leu Val Ile Phe Pro Glu Gly Thr Arg
 165 170 175
 Tyr Asn Pro Glu Gln Thr Lys Val Leu Ser Ala Ser Gln Ala Phe Ala
 180 185 190
 Ala Gln Arg Gly Leu Ala Val Leu Lys His Val Leu Thr Pro Arg Ile
 195 200 205
 Lys Ala Thr His Val Ala Phe Asp Cys Met Lys Asn Tyr Leu Asp Ala
 210 215 220
 Ile Tyr Asp Val Thr Val Val Tyr Glu Gly Lys Asp Asp Gly Gly Gln
 225 230 235 240
 Arg Arg Glu Ser Pro Thr Met Thr Glu Phe Leu Cys Lys Glu Cys Pro
 245 250 255
 Lys Ile His Ile His Ile Asp Arg Ile Asp Lys Lys Asp Val Pro Glu
 260 265 270
 Glu Gln Glu His Met Arg Arg Trp Leu His Glu Arg Phe Glu Ile Lys
 275 280 285
 Asp Lys Met Leu Ile Glu Phe Tyr Glu Ser Pro Asp Pro Glu Arg Arg
 290 295 300
 Lys Arg Phe Pro Gly Lys Ser Val Asn Ser Lys Leu Ser Ile Lys Lys
 305 310 315 320
 Thr Leu Pro Ser Met Leu Ile Leu Ser Gly Leu Thr Ala Gly Met Leu
 325 330 335
 Met Thr Asp Ala Gly Arg Lys Leu Tyr Val Asn Thr Trp Ile Tyr Gly
 340 345 350
 Thr Leu Leu Gly Cys Leu Trp Val Thr Ile Lys Ala
 355 360

<210> 6
 <211> 26
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> misc_binding
 <222> 1..26
 <223> primer oligonucleotide GC1.5p.1
 <400> 6
 ctgtccctgg tgctccacac gtactc
 <210> 7
 <211> 26
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> misc_binding
 <222> 1..26
 <223> primer oligonucleotide GC1.5p.2
 <400> 7
 tgggtgctcca cacgtactcc atgcgc

26

26

```

<210> 8
<211> 27
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..27
<223> primer oligonucleotide pg15RACE196
<400> 8
caatatctgg accccggtgt aattctc
27
<210> 9
<211> 34
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..34
<223> primer oligonucleotide GC1.3p
<400> 9
cttgctgct ggagacacag aatttcgata gcac
34
<210> 10
<211> 24
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..24
<223> primer oligonucleotide PGRT32
<400> 10
tttttttttt tttttttttg aaat
24
<210> 11
<211> 6
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> 160..165
<223> box2 from SEQID4, present in AF003136, P33333, P26647, U89336,
U56417, AB005623.
<400> 11
Phe Pro Glu Gly Thr Arg
1 5
<210> 12
<211> 6
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> 129..134
<223> box2 from Z72511
<400> 12
Phe Pro Glu Gly Thr Asp
1 5
<210> 13
<211> 6
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> 223..228
<223> box2 from P38226, Z49770

```

<400> 13
Phe Pro Glu Gly Thr Asn
1 5
<210> 14
<211> 6
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> 90..95
<223> box2 from Z49860 and Z29518
<400> 14
Phe Val Glu Gly Thr Arg
1 5
<210> 15
<211> 9
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> 211..219
<223> box3 from SEQID4, present in AF003136
<400> 15
Leu Asp Ala Ile Tyr Asp Val Thr Val
1 5
<210> 16
<211> 9
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> 204..212
<223> box3 from Z72511
<400> 16
Val Glu Tyr Ile Tyr Asp Ile Thr Ile
1 5
<210> 17
<211> 9
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> 271..279
<223> box3 from P38226
<400> 17
Ile Glu Ser Leu Tyr Asp Ile Thr Ile
1 5
<210> 18
<211> 9
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> 265..273
<223> box3 from Z49770
<400> 18
Leu Asp Ala Ile Tyr Asp Val Thr Ile
1 5
<210> 19
<211> 9
<212> PRT

```

<213> Homo sapiens
<220>
<221> SITE
<222> 138..146
<223> box3 fromZ49860
<400> 19
Val Pro Ala Ile Tyr Asp Met Thr Val
1 5
<210> 20
<211> 9
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> 218..226
<223> box3 from Z29518
<400> 20
Val Pro Ala Ile Tyr Asp Thr Thr Val
1 5
<210> 21
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-123
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-123.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-123.mis2
<400> 21
tttctcatcc tcacacctca ctgcgccct cctgaaccca ctccttt
<210> 22
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-26
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-26.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-26.mis2
<400> 22
ccctgtnaga cacgtcctgt atcgttggtg agatgggaaa gtgcatc
<210> 23
<211> 47
<212> DNA
<213> Homo Sapiens

```

47

47

<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-14
<221> allele
<222> 24
<223> polymorphic base T
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-14.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-14.mis2
<400> 23
gcaggagaca gaccagacat gatttggttct agtctagctg attcata 47
<210> 24
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-77, extracted from SEQ ID1 12057 12103
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-77.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-77.mis2
<400> 24
gctgttcaga cttaaacttg agactacagt cagtcagaga acttgct 47
<210> 25
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-217, extracted from SEQ ID1 34469 34515
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-217.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-217.mis2
<400> 25
atatagttca cgttatgttc atacttaatt gttgcatttt gtttgcc 47
<210> 26
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-67, extracted from SEQ ID1 51612 51658

<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-67.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-67.mis2
<400> 26
gccagtgaag tacagactta attcgatcatg actgaacgaa ttgtgtt 47
<210> 27
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-213
<221> allele
<222> 24
<223> polymorphic base T
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-213.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-213.mis2
<400> 27
ccttagcatt caagcccctg agctctggtg ttgtccaccc ctggggg 47
<210> 28
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-221
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-221.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-221.mis2
<400> 28
agcttgagaa accagaaaag ccaaaaggag gctcctacca catgggt 47
<210> 29
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-135
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind

<222> 1..23
<223> potential microsequencing oligo 99-135.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-135.mis2
<400> 29
agtcactata tctatgttta atgaagatag aaagagatgc agaaatg 47
<210> 30
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-123, variant version of SEQ ID21
<221> allele
<222> 24
<223> base T ; C in SEQ ID21
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-123.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-123.mis2
<400> 30
tttctcatcc tcacacctca ctgtgccct cctgaacca ctccttt 47
<210> 31
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-26, variant version of SEQ ID22
<221> allele
<222> 24
<223> base A ; G in SEQ ID22
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-26.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-26.mis2
<400> 31
ccctgttnaga cacgtcctgt atcattgttg agatgggaaa gtgcatc 47
<210> 32
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-14, variant version of SEQ ID23
<221> allele
<222> 24
<223> base C ; T in SEQ ID23
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-14.mis1
<221> primer_bind
<222> 25..47

<223> complement potential microsequencing oligo 4-14.mis2
<400> 32
gcaggagca gaccagacat gatctgttct agtctagctg attcata
<210> 33
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-77, variant version of SEQ ID24
<221> allele
<222> 24
<223> base G ; C in SEQ ID24
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-77.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-77.mis2
<400> 33
gctgttcaga ctaaacttgg agagtacagt cagtcagaga acttgct
<210> 34
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-217, variant version of SEQ ID25
<221> allele
<222> 24
<223> base T ; C in SEQ ID25
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-217.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-217.mis2
<400> 34
atatagttca cgttatgttc atatttaatt gttgcatttt gtttgcc
<210> 35
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-67, variant version of SEQ ID26
<221> allele
<222> 24
<223> base T ; C in SEQ ID26
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-67.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-67.mis2
<400> 35
gccagtgaat tacagactta atttgtcatg actgaacgaa tttgttt
<210> 36

47

47

47

47

<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-213, variant version of SEQ ID27
<221> allele
<222> 24
<223> base C ; T in SEQ ID27
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-213.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-213.mis2
<400> 36
ccttagcatt caagccctg agccctggtg ttgtccaccc ctggggg 47
<210> 37
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-221, variant version of SEQ ID28
<221> allele
<222> 24
<223> base C ; A in SEQ ID28
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-221.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-221.mis2
<400> 37
agcttgagaa accagaaaag ccacaaggag gctcctacca catgggt 47
<210> 38
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-135, variant version of SEQ ID29
<221> allele
<222> 24
<223> base G ; A in SEQ ID29
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-135.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-135.mis2
<400> 38
agtactata tctatgttta atggagatag aaagatatgc agaaatg 47
<210> 39
<211> 18
<212> DNA
<213> Homo Sapiens
<220>

```

<221> primer_bind
<222> 1..18
<223> upstream amplification primer 99-123-PU
<400> 39
aaagccagga ctagaagg
<210> 40
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer 4-26-PU
<400> 40
tacagccctg taagacac
<210> 41
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer 4-14-PU
<400> 41
tctaacctct catccaac
<210> 42
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer 4-77-PU, extracted from SEQ ID1 11930
11947
<400> 42
tgttgattta caggcggc
<210> 43
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> upstream amplification primer 99-217-PU, extracted from SEQ ID1 34216
34234
<400> 43
ggtgggaatt tactatatg
<210> 44
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer 4-67-PU, extracted from SEQ ID1 51596
51613
<400> 44
aagttcacct tctcaagc
<210> 45
<211> 20
<212> DNA

```

```

<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> upstream amplification primer 99-213-PU
<400> 45
atactggcag cgtgtgcttc
<210> 46
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> upstream amplification primer 99-221-PU
<400> 46
ccctttttct tcaactgttc
<210> 47
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer 99-135-PU
<400> 47
tggaagtgtg tattgccc
<210> 48
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer 99-123-RP
<400> 48
tattcagaaa ggagtggg
<210> 49
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer 4-26-RP
<400> 49
tgaggactgc taggaaag
<210> 50
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> downstream amplification primer 4-14-RP
<400> 50
gactgtatcc tttgatgcac
<210> 51
<211> 20
<212> DNA
<213> Homo Sapiens

```

```

<220>
<221> primer_bind
<222> 1..20
<223> downstream amplification primer 4-77-RP, extracted from SEQ ID1 12339
123
58 complement
<400> 51
ggaaaggtag tcattcatag                20
<210> 52
<211> 21
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..21
<223> downstream amplification primer 99-217-RP, extracted from SEQ ID1 34625
34645 complement
<400> 52
gtttattttg tgtgagcttt g                21
<210> 53
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> downstream amplification primer 4-67-RP, extracted from SEQ ID1 51996
520
15 complement
<400> 53
tgaaagagtt tattctctgg                20
<210> 54
<211> 21
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..21
<223> downstream amplification primer 99-213-RP
<400> 54
ttattgcccc acatgcttga g                21
<210> 55
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> downstream amplification primer 99-221-RP
<400> 55
tcattcgtct ggctaggtc                19
<210> 56
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer 99-135-RP
<400> 56
aaacacctcc cattgtgc                18

```

<210> 57
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1482
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1482.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1482.mis2
<400> 57
agtgaagtct gagggggaaa aatcaaccct atagagggaa ggatctg 47
<210> 58
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-73, extracted from SED ID1 13657 13703
<221> allele
<222> 24
<223> polymorphic base C in PG1 (13680) SEQ ID1
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-73.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-73.mis2
<400> 58
gttttcctta tgatgttaca tggcttattt tttaaaggtaa tgaaaac 47
<210> 59
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-65, extracted from SEQ ID1 51448 51494
<221> allele
<222> 24
<223> polymorphic base T in PG1 (51471) SEQ ID1
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-65.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-65.mis2
<400> 59
ggtgctgctc agcggcttgc acgtagactt gctaggaaga aatgcag 47
<210> 60
<211> 47
<212> DNA
<213> Homo Sapiens

<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1482, variant version of SEQ ID57
<221> allele
<222> 24
<223> base A ; C in SEQ ID57
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1482.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1482.mis2
<400> 60
agtgaagtct gagggggaaa aataaacct atagaggga ggatctg
<210> 61
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-73, variant version of SEQ ID58
<221> allele
<222> 24
<223> base G ; C in SEQ ID58
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-73.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-73.mis2
<400> 61
gttttcctta tgatgttaca tgggttattt ttaaaggtaa tgaaaac
<210> 62
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-65, variant version of SEQ ID59
<221> allele
<222> 24
<223> base C ; T in SEQ ID59
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-65.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-65.mis2
<400> 62
ggtgctgctc agcggcttgc acgcagactt gctaggaaga aatgcag
<210> 63
<211> 21
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..21
<223> upstream amplification primer 99-1482-PU

47

47

47

<400> 63
atcaaatacag tgaagtctga g 21
<210> 64
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer 4-73-PU, extracted from SEQ ID1 13547
13564
<400> 64
atcgctggaa cattctgg 18
<210> 65
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> upstream amplification primer 4-65-PU, extracted from SEQ ID1 51149
51168
<400> 65
gatttaagct acgctattag 20
<210> 66
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> downstream amplification primer 99-1482-RP
<400> 66
acaaatctat ataaggctgg 20
<210> 67
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> downstream amplification primer 4-73-RP, extracted from SEQ ID1 13962
13981 complement
<400> 67
ctcttggtta aacagcagtg 20
<210> 68
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer 4-65-RP, extracted from SEQ ID1 51482
51499 complement
<400> 68
tggtctgca tttcttcc 18
<210> 69
<211> 5226
<212> DNA
<213> Homo sapiens
<400> 69

ctgctgtccc	tggtgctcca	cacgtactcc	atg	cgc	tac	ctg	ctg	ccc	agc	gtc	54
			Met	Arg	Tyr	Leu	Leu	Pro	Ser	Val	
			1				5				
gtg	ctc	ctg	ggc	acg	gcg	ccc	acc	tac	gtg	ttg	102
Val	Leu	Leu	Gly	Thr	Ala	Pro	Thr	Tyr	Val	Leu	
10						15				20	
cgg	ctg	ctc	tcc	gcc	ttc	ctg	ccc	gcc	cgc	ttc	150
Arg	Leu	Leu	Ser	Ala	Phe	Leu	Pro	Ala	Arg	Phe	
25					30				35		40
gac	cgg	ctc	tac	tgc	gtc	tac	cag	agc	atg	gtg	198
Asp	Arg	Leu	Tyr	Cys	Val	Tyr	Gln	Ser	Met	Val	
				45					50		55
aat	tac	acc	ggg	gtc	cag	ata	ttg	cta	tat	gga	246
Asn	Tyr	Thr	Gly	Val	Gln	Ile	Leu	Leu	Tyr	Gly	
			60					65			70
aaa	gaa	aat	ata	ata	tat	tta	gca	aat	cat	caa	294
Lys	Glu	Asn	Ile	Ile	Tyr	Leu	Ala	Asn	His	Gln	
		75				80				85	
att	gtt	gct	gac	atc	ttg	gcc	atc	agg	cag	aat	342
Ile	Val	Ala	Asp	Ile	Leu	Ala	Ile	Arg	Gln	Asn	
90						95				100	
cgc	tac	gtg	ctg	aaa	gaa	ggg	tta	aaa	tgg	ctg	390
Arg	Tyr	Val	Leu	Lys	Glu	Gly	Leu	Lys	Trp	Leu	
105					110				115		120
tac	ttt	gct	cag	cat	gga	gga	atc	tat	gta	aag	438
Tyr	Phe	Ala	Gln	His	Gly	Gly	Ile	Tyr	Val	Lys	
				125					130		135
aac	gag	aaa	gag	atg	cga	aac	aag	ttg	cag	agc	486
Asn	Glu	Lys	Glu	Met	Arg	Asn	Lys	Leu	Gln	Ser	
			140					145			150
act	cca	atg	tat	ctt	gtg	att	ttt	cca	gaa	ggg	534
Thr	Pro	Met	Tyr	Leu	Val	Ile	Phe	Pro	Glu	Gly	
		155					160				165
gag	caa	aca	aaa	gtc	ctt	tca	gct	agt	cag	gca	582
Glu	Gln	Thr	Lys	Val	Leu	Ser	Ala	Ser	Gln	Ala	
170						175				180	
ggc	ctt	gca	gta	tta	aaa	cat	gtg	cta	aca	cca	630
Gly	Leu	Ala	Val	Leu	Lys	His	Val	Leu	Thr	Pro	
185					190					195	200
cac	gtt	gct	ttt	gat	tgc	atg	aag	aat	tat	tta	678
His	Val	Ala	Phe	Asp	Cys	Met	Lys	Asn	Tyr	Leu	
				205					210		215
gtt	acg	gtg	gtt	tat	gaa	ggg	aaa	gac	gat	gga	727
Val	Thr	Val	Val	Tyr	Glu	Gly	Lys	Asp	Asp	Gly	
				220				225			
caccgaccat	gacggaat	ctctgcaa	aatgtccaaa	aattcatatt	cacattgac	787					
gtatcgacaa	aaaagatg	ccagaaga	aagaacatat	gagaagatg	ctgcatga	847					
gtttcgaaat	caaagata	atgcttat	aattttatga	gtcaccagat	ccagaaaaga	907					
gaaaaagatt	tcctgggaaa	agtgttaatt	ccaaattaag	tatcaagaag	actttaccat	967					
caatgttgat	cttaagtgg	ttgactgc	gcatgcttat	gaccgatg	ggaaggagc	1027					
tgtatgtgaa	cacctggata	tatggaac	tacttggtg	cctgtgggt	actattaaag	1087					
catagacaag	tagctgtctc	cagacagt	gatgtgctac	attgtctatt	tttggcggt	1147					
gcacatgaca	tcaaattgtt	tcctgaatt	attaaggagt	gtaaataaag	ccttggtgat	1207					
tgaagattgg	ataataga	ttgtgacga	agctgatatg	caatggctt	gggcaaacat	1267					
acctggttgt	acaactttag	catcgggg	gctggaagg	taaaagctaa	atggagtctc	1327					
tcctgctctg	tccatttcct	atgaactaat	gacaacttga	gaaggctggg	aggatttgt	1387					
atcttgcaag	tcagattgg	gcatttttga	gcattaattt	gcagcgtatt	tcactttttc	1447					
tggtattttc	aattttattac	aacttgacag	ctccaagctc	ttattactaa	agtatttagt	1507					
atcttgacag	tagttaatat	ttcatctttt	gcttattttc	acaagtcagt	gaaataaatt	1567					
gtatttagga	agtgtcagga	tggtcaaaag	aaagggtaaa	aagtgttc	ggggaaaaag	1627					
ctctgttttag	cacatgattt	tattgtattg	cgttatttag	tgattttact	cattttatat	1687					

ttgcaaaata	aattttcta	atttattgaa	attgctta	ttgcacacc	tgtacacaca	1747
gaaaaatg	taaaatat	gaacgaagt	taaaattgt	actctgatt	attatagcag	1807
aacttttaa	ttcccagct	tttgaagatt	taagctacgc	tattagtact	tccctttgtc	1867
tgtgccata	gtgcttgaa	acgttaaggt	tttctgtttt	gttttgtttt	tttaatatca	1927
aaagagtcg	gttgaacct	ggttggacc	caagttcaca	agatttttaa	ggtgatgaga	1987
gcctgcagac	attctgccta	gatttactag	cgtgtgcctt	ttgcctgctt	ctcttttgatt	2047
tcacagaata	ttcattcaga	agtcgcgttt	ctgtagtgtg	gtggattccc	actgggctct	2107
ggctccttccc	ttggatcccg	tcagtgggtg	tgctcagcgg	cttgacagta	gacttgctag	2167
gaagaaatgc	agagccagcc	tgtgctgccc	actttcagag	ttgaactcct	taagcccttg	2227
tgagtgggct	tcaccagcta	ctgcagaggc	attttgcatt	tgtctgtgtc	aagaagttca	2287
ccttctcaag	ccagtgaat	acagacttaa	ttcgtcatga	ctgaacgaat	ttgtttattt	2347
cccatagggt	ttagtggagc	tacacattaa	tatgtatcgc	cttagagcaa	gagctgtgtt	2407
ccaggaacca	gatcacgatt	tttagccatg	gaacaatata	tcccatggga	gaagaccttt	2467
cagtgtgaac	tgttctattt	ttgtgttata	atttaaacct	cgatttccct	atagtccttt	2527
aagtgtacat	ttctgtctac	tgctactgga	tttttgctgc	agaaatatat	cagtggccca	2587
cattaaacat	accagttgga	tcagtataag	caaaatgaaa	gaaataatga	ttaagggaaa	2647
attaagtgc	tggtgtacac	tgcttctccc	atgccagaga	ataaactcct	tcaagcatca	2707
tctttgaaga	gtcgtgtggt	gtgaattggg	ttgtgtacat	tagaatgtat	gcacacatcc	2767
atggacactc	aggatatagt	tgccctaata	atcggggcat	gggtaaaact	tatgaaaatt	2827
tcctcatgct	gaattgtaat	ttctctctac	ctgtaaagta	aaatttagat	caattccatg	2887
tctttgttaa	gtacagggat	ttaatatatt	ttgaatataa	tggttatgtt	ctaaatttga	2947
actttgagag	gcaatactgt	tggaattatg	tggaattctaa	ctcattttaa	caaggtagcc	3007
tgacctgcat	aagatcactt	gaatgttagg	tttcatagaa	ctatactaat	cttctcacia	3067
aaggtctata	aaatacagtc	gttgaaaaaa	attttgtatc	aaaatgtttg	gaaaattaga	3127
agcttctcct	taacctgtat	tgatactgac	ttgaattatt	ttctaaaatt	aagagccgta	3187
tacctacctg	taagtctttt	cacatatcat	ttaaactttt	gtttgtatta	ttactgtatt	3247
acagcttagt	tattaatttt	tctttataag	aatgccgtcg	atgtgcatgc	ttttatgttt	3307
ttcagaaaag	ggtgtgtttg	gatgaaagta	aaaaaaaaaa	taaaatcctt	cactgtctct	3367
aatggctgtg	ctgttttaaca	ttttttgacc	ctaaaattca	ccaacagctc	ccagtcacat	3427
aaaatagggt	taatgactgg	ccctgcattc	ttcacaaat	ttttccctaa	gctttgagca	3487
aagtttttaa	aaaatacact	aaaataatca	aaactgttaa	gcagtatat	agtttgggta	3547
tataaattca	tctgcaattt	ataagatgca	tgcccgatgt	taatttgctt	ggcaattctg	3607
taatcattaa	gtgatctcag	tgaacaatgt	caaatgcctt	aaatttaact	aagttgggta	3667
taaaagtgcc	gatctggcta	actcttacac	catacatact	gatagttttt	catatgtttc	3727
atttccatgt	gattttttaa	atttagagtg	gcaacaattt	tgcttaatat	gggttacata	3787
agctttattt	tttcccttgt	tcataattat	attctttgaa	taggtctgtg	tcaatcaagt	3847
gatcttaact	gactgatcat	agatagaagg	aaataaggcc	aagttcaaga	ccagcctggg	3907
caacatatcg	agaacctgtc	tacaaaaaaa	ttaaaaaaa	ttagccaggc	atggtggcgt	3967
acactgagta	gtttgtccca	gctactcggg	agggtgaggt	gggaggatcg	cttcagccca	4027
ggagggttag	attgcagtga	gccatggaca	taccactgca	ctacagccta	ggtaacagca	4087
cgagacccca	actcttagaa	aatgaaaagg	aaatatagaa	atataaaatt	tgcttattat	4147
agacacacag	taactcccag	atatgtacca	caaaaaatgt	gaaaagagag	agaaatgtct	4207
accaaagcag	tattttgtgt	gtataattgc	aagcgcatag	taaaataatt	ttaaccttaa	4267
tttgttttta	gtagtgttta	gattgaagat	tgagtgaat	attttcttgg	cagatatctc	4327
gtatctgggt	gaaagctaca	atgcaatgtc	gttgtagttt	tgcatggcct	gctttataaa	4387
caagattttt	tctccctcct	tttgggccag	ttttcattac	gagtaactca	cactttttga	4447
ttaaagaact	tgaattacg	ttatcactta	gtataattga	cattatatag	agactatgta	4507
acatgcaatc	attagaatca	aaattagtag	tttgggtcaaa	atattttaca	cattcacata	4567
cttgtcaaat	attcatgtaa	ttactgaat	ttaaaacctt	caactattat	gaagtgtctg	4627
tctgtacaat	cgctaattta	ctcagtttag	agtagctaca	actcttcgat	actatcatca	4687
atatttgaca	tcttttccaa	tttgtgtatg	aaaagtaaat	ctattcctgt	agcaactggg	4747
gagtcataata	tgagggtcaaa	gacatatacc	ttgttattat	aatatgtata	ctataataat	4807
agctgggttat	cctgagcagg	ggaaaagggt	atttttagga	aaaccacttc	aaatagaaa	4867
ctgaagtact	tctaataatc	tgagggaagt	ataatatgtg	gaacaaactc	tcaacaaaat	4927
gtttattgat	gttgatgaaa	cagatcagtt	tttccatccg	gattattatt	ggttcatgat	4987
tttatatgtg	aatatgtaag	atatgttctg	caattttata	aatgttcatg	tcttttttta	5047
aaaaagggtgc	tatcgaaatt	ctgtgtctcc	agcaggcaag	aatacttgac	taactccttt	5107
tgtctctttta	tggtattttc	agaataaagt	ctgacttgtg	tttttgagat	tattgggtgcc	5167
tcattaattc	agcaataaag	gaaaatatgc	atttcaaaaa	naaaaaaaa	aaaaaaaaa	5226

<210> 70
<211> 228

<212> PRT

<213> Homo sapiens

<400> 70

```

Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
1      5      10      15
Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
20      25      30
Ala Arg Phe Tyr Gln Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln
35      40      45
Ser Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln Ile Leu
50      55      60
Leu Tyr Gly Asp Leu Pro Lys Asn Lys Glu Asn Ile Ile Tyr Leu Ala
65      70      75      80
Asn His Gln Ser Thr Val Asp Trp Ile Val Ala Asp Ile Leu Ala Ile
85      90      95
Arg Gln Asn Ala Leu Gly His Val Arg Tyr Val Leu Lys Glu Gly Leu
100     105     110
Lys Trp Leu Pro Leu Tyr Gly Cys Tyr Phe Ala Gln His Gly Gly Ile
115     120     125
Tyr Val Lys Arg Ser Ala Lys Phe Asn Glu Lys Glu Met Arg Asn Lys
130     135     140
Leu Gln Ser Tyr Val Asp Ala Gly Thr Pro Met Tyr Leu Val Ile Phe
145     150     155     160
Pro Glu Gly Thr Arg Tyr Asn Pro Glu Gln Thr Lys Val Leu Ser Ala
165     170     175
Ser Gln Ala Phe Ala Ala Gln Arg Gly Leu Ala Val Leu Lys His Val
180     185     190
Leu Thr Pro Arg Ile Lys Ala Thr His Val Ala Phe Asp Cys Met Lys
195     200     205
Asn Tyr Leu Asp Ala Ile Tyr Asp Val Thr Val Val Tyr Glu Gly Lys
210     215     220
Asp Asp Gly Gly
225

```

<210> 71

<211> 158

<212> DNA

<213> Homo sapiens

<400> 71

```

gccttgcaagt attaaaacat gtgctaacac cacgaataaa ggcaactcac gttgcttttg      60
attgcatgaa gaattattta gatgcaattt atgatgttac ggtgggttat gaagggaaag      120
acgatggagg gtagcgaaga gagtaccga ccatgacg      158

```

<210> 72

<211> 1381

<212> DNA

<213> Mus musculus

<220>

<221> misc_binding

<222> 608..629

<223> amplification primer g34292.pu

<221> misc_binding

<222> 740..758

<223> amplification primer g34292.rp

<400> 72

```

gagccgagag gatgctgctg tccctggtgc tccacacgta ctct atg cgc tac ctg      56
                               Met Arg Tyr Leu
                               1
ctc ccc agc gtc ctg ttg ctg ggc tgc gcg ccc acc tac ctg ctg gcc      104
Leu Pro Ser Val Leu Leu Leu Gly Ser Ala Pro Thr Tyr Leu Leu Ala
5      10      15      20
tgg acg ctg tgg cgg gtg ctc tcc gcg ctg atg ccc gcc cgc ctg tac      152
Trp Thr Leu Trp Arg Val Leu Ser Ala Leu Met Pro Ala Arg Leu Tyr

```

																25																	30																	35																	
cag	cg	gtg	gac	gac	cg	ctt	tac	tgc	gtc	tac	cag	aac	atg	gtg	ctc		200																																																		
Gln	Arg	Val	Asp	Asp	Arg	Leu	Tyr	Cys	Val	Tyr	Gln	Asn	Met	Val	Leu																																																				
																40																	45																	50																	
ttc	ttc	ttc	gag	aac	tac	acc	ggg	gtc	cag	ata	ttg	cta	tat	gga	gat		248																																																		
Phe	Phe	Phe	Glu	Asn	Tyr	Thr	Gly	Val	Gln	Ile	Leu	Leu	Tyr	Gly	Asp																																																				
																55																	60																	65																	
ttg	cca	aaa	aat	aaa	gaa	aat	gta	ata	tat	cta	gcg	aat	cat	caa	agc		296																																																		
Leu	Pro	Lys	Asn	Lys	Glu	Asn	Val	Ile	Tyr	Leu	Ala	Asn	His	Gln	Ser																																																				
																70																	75																	80																	
aca	gtt	gac	tgg	att	gtt	gcg	gac	atg	ctg	gct	gcc	aga	cag	gat	gcc		344																																																		
Thr	Val	Asp	Trp	Ile	Val	Ala	Asp	Met	Leu	Ala	Ala	Arg	Gln	Asp	Ala																																																				
																85																	90																	95																	100
cta	gga	cat	gtg	cg	tac	gta	ctg	aaa	gac	aag	tta	aaa	tgg	ctt	ccg		392																																																		
Leu	Gly	His	Val	Arg	Tyr	Val	Leu	Lys	Asp	Lys	Leu	Lys	Trp	Leu	Pro																																																				
																105																	110																	115																	
ctg	tat	ggg	ttc	tac	ttt	gct	cag	cat	gga	gga	att	tat	gta	aaa	cga		440																																																		
Leu	Tyr	Gly	Phe	Tyr	Phe	Ala	Gln	His	Gly	Gly	Ile	Tyr	Val	Lys	Arg																																																				
																120																	125																	130																	
agt	gcc	aaa	ttt	aat	gat	aaa	gaa	atg	aga	agc	aag	ctg	cag	agc	tat		488																																																		
Ser	Ala	Lys	Phe	Asn	Asp	Lys	Glu	Met	Arg	Ser	Lys	Leu	Gln	Ser	Tyr																																																				
																135																	140																	145																	
gtg	aac	gca	gga	aca	ccg	atg	tat	ctt	gtg	att	ttc	cca	gag	gga	aca		536																																																		
Val	Asn	Ala	Gly	Thr	Pro	Met	Tyr	Leu	Val	Ile	Phe	Pro	Glu	Gly	Thr																																																				
																150																	155																	160																	
agg	tat	aat	gca	aca	tac	aca	aaa	ctc	ctt	tca	gcc	agt	cag	gca	ttt		584																																																		
Arg	Tyr	Asn	Ala	Thr	Tyr	Thr	Lys	Leu	Leu	Ser	Ala	Ser	Gln	Ala	Phe																																																				
																165																	170																	175																	180
gct	gct	cag	cg	ggc	ctt	gca	gta	tta	aaa	cac	gta	ctg	aca	cca	aga		632																																																		
Ala	Ala	Gln	Arg	Gly	Leu	Ala	Val	Leu	Lys	His	Val	Leu	Thr	Pro	Arg																																																				
																185																	190																	195																	
ata	aag	gcc	act	cac	gtt	gct	ttt	gat	tct	atg	aag	agt	cat	tta	gat		680																																																		
Ile	Lys	Ala	Thr	His	Val	Ala	Phe	Asp	Ser	Met	Lys	Ser	His	Leu	Asp																																																				
																200																	205																	210																	
gca	att	tat	gat	gtc	aca	gtg	gtt	tat	gaa	ggg	aat	gag	aaa	ggt	tca		728																																																		
Ala	Ile	Tyr	Asp	Val	Thr	Val	Val	Tyr	Glu	Gly	Asn	Glu	Lys	Gly	Ser																																																				
																215																	220																	225																	
gga	aaa	tac	tca	aat	cca	cca	tcc	atg	act	gag	ttt	ctc	tgc	aaa	cag		776																																																		
Gly	Lys	Tyr	Ser	Asn	Pro	Pro	Ser	Met	Thr	Glu	Phe	Leu	Cys	Lys	Gln																																																				
																230																	235																	240																	
tgc	cca	aaa	ctt	cat	att	cac	ttt	gat	cgt	ata	gac	aga	aat	gaa	gtt		824																																																		
Cys	Pro	Lys	Leu	His	Ile	His	Phe	Asp	Arg	Ile	Asp																																																								

```

gcaagtagca ggctgcagtc acagtctctt attgatggct acacattgta tcacattggt 1169
tcctgaatta aataaggagt tttcttggtg ttgttttttt tgttttggtt tggtctggtt 1229
taagccttga tgatngnnn cnnnnnnnnn ncnantcnng ngaccacagc caacatgcat 1289
ttgatttggg gcaaacacat gtggcttttc aggtgctggg gttgctggag acatggaagc 1349
taagtggagt ttatgctgtt tttttttttt tt 1381
<210> 73
<211> 15766
<212> DNA
<213> Mus musculus
<220>
<221> exon
<222> 52..121
<223> exon2
<221> exon
<222> 682..797
<223> exon3
<221> exon
<222> 2628..2717
<223> exon4
<221> exon
<222> 7834..7924
<223> exon5
<221> exon
<222> 9804..9965
<223> exon6
<221> exon
<222> 11404..11527
<223> exon7
<221> exon
<222> 13539..14035
<223> exon8
<221> misc_feature
<222> 13762..13764
<223> stop CDS
<221> polyA_signal
<222> 13835..13839
<223> AATAAA potential
<400> 73
tttttttttt ttaattgtca aagtcatgat tctttttggt ttctctttta gatattgcta 60
tatggagatt tgccaaaaaa taaagaaaat gtaatatatc tagcgaatca tcaaagcaca 120
ggtttgtatt tcatttgatg aaatttgggt ttttctagaa atggtaaagt agcattaata 180
tgtacacaca catacacaca aacacacata tgtacacaca catatgtttt aaagacagga 240
tttcatgtga cccagaatgg cctcactctc tctgagtagc tgagaatgat ttttaagctg 300
tgacacacct gccttcactc ccaaggtaga ggaattgcag gtgctttctt tgnnnnnnnn 360
nnnnnnnnnn tttgagtttt ggggaggggg tatatttttt aatgtgtctg tagttggctt 420
tgttttaagc attttaatca tactttattt ttaaaaaaac taaaagcttt ttttaaggct 480
ggctcttgcta tgtggcccta gtgttcctgg gacttgctct gtacaccggg ttgactctga 540
gcctgtgcgc cttctgcctc tgcctccata gttagattct caggacatgt taaaaagact 600
gtgctgtgaa gatgagtttt tgttcctggg agggaaaggt ggagctgact tgtgaggtac 660
tgacttgggt ctgccttaca gttgactgga ttgttgcgga catgctggct gccagacagg 720
atgccctagg acatgtgcgc tacgtactga aagacaagtt aaaatggctt ccgctgtatg 780
ggttctactt tgctcaggta aactttgtct ttgccccttt atttcaaact taacaccatt 840
taatgaaact atatctgatt tttttgttta tgtgtttggt ttatgggtacc cgtgattgaa 900
catggggtca tatgtgtgct actgagtgac agccttagtt cagacatttt ttaaagcgac 960
ttttactagt atttttattt agaattctat atgtgtgcac atgcatatgt gtgcttggtg 1020
gcacacgtgg atgcatgtga ggtcgaagga caattttcag tacaagtgtg agtgtcactt 1080
tttaggcacc ttccactctt attttgagac agtctcctag acctttgctg agttgcccag 1140
gctagccggc cagtgagecc tgggcatcta ccggtctctg cctccttacc tttacttagg 1200
ttacaagtgt gtgctgtcac gccagctgt ttactagatt ctagggatcc aaatgtgggt 1260
cctcgtaact tgtgagacaa gtactttcca aactgagcca cctccctagc tcttcttcac 1320
ggttcctgat ggtgtgtgtc tagatggctg gttgtccgta tatttaagtc cagtagcaga 1380

```

aatacaata	cctaggagtc	caatagaaag	ctacaagtc	agaattgaca	atcggtaatg	1440
ttcggaaatt	gattcaaaag	tagtttagtga	gtgacagaca	ggagctaaaa	gcagactctg	1500
agctcagagt	gtgaagtgtg	gagaaatgtg	ttttctcaca	gttctgaagg	ctgaaagtct	1560
ccccaaagtc	aggatgtggg	tggtactgct	gtctcccaan	caccacctc	tttggattat	1620
agactgcagc	cttctccctg	tggtctgagc	cggectttcc	cacatgtgga	catccttggg	1680
gggtgttcca	ccagcagggc	ctcagctagt	gccccttatt	cacttaactg	taattgattt	1740
cttaaaagacc	ctgtctccat	acacagtcac	tggtgaagct	gaagcttcaa	tgtaagagtt	1800
aaggggggag	ggggaaattt	agtccataat	gggtgcacac	caatctctgt	agctgagttc	1860
atgattcagt	tctttaaagg	ctctgagtgt	agacattatc	ttaattattt	tgcccattta	1920
tgtattatct	tttaatttatt	ttatgtaaat	gaatgcctgt	gtatatatgt	ttctgggtcc	1980
tagtccatat	tttaattcct	taaaggatgg	agggtagac	ttttgtcttt	tttaatttct	2040
atccttccct	cctggcctcc	tggtggcctct	tttacgtatt	tattattttt	aatttatttt	2100
atgtgtttga	gtgtttttga	tctatgcatt	cctggggccc	atggaggtca	gaaaaacaca	2160
ttaggtggcc	tacaactgag	ttatgggtgg	tttgtgacca	tggtgtgctg	ggacttgatc	2220
accagtctct	gagaagactc	gtgtctgctg	agccttctct	ccagctctgg	gagtggtgat	2280
attttaagga	tacttttaat	tgacttggtg	aatgacagta	gaaaatcaat	gagtaggat	2340
ccatcggaaa	aagcttttga	actaaatctt	ttaaagagaa	aatattttta	gtgctaacaa	2400
aattaaatgt	gtattttcca	tgatgcagtt	ttacttgggc	tctgtagaaa	taggattttc	2460
aggtagatat	tgtatatata	gttggaata	tttaataact	aactgtcgct	tgagtctcga	2520
aatgtagttt	tatgtttttt	actcattagg	agtacagttg	ccttaataac	tacggagatt	2580
agttattaaa	gaataattgc	tcttcttttt	tcttttctgt	gtaccagcat	ggaggatttt	2640
atgtaaaacg	aagtgcctaa	tttaatgata	aagaaatgag	aagcaagctg	cagagctatg	2700
tgaacgcagg	aacaccggta	agtgcgccc	cttttattcc	tcaaggcagg	ttaaagattt	2760
aagtctctaa	gtcattttga	aaatatatta	ccccatgtgg	agcaatggaa	ctggttcggg	2820
gttttgggtg	gataagctgt	cctctggccg	tgaggtaaga	ttgctgcagg	tgattgttaag	2880
gtttctcctg	agtaacagtc	agcatgggct	cgggacgggc	aagggcaggc	cttagtgtgc	2940
agaggatgga	gctcactgaa	gccccaaaga	gttagtcttc	acatgagatt	cagttctaga	3000
agaagttaaa	ttgctttctt	tctgtgtaaa	tttggatttt	tattgtagaa	attaaagt	3060
gttttctttt	aaaacaaaca	caaaccaga	gcaaagagtc	tcctagtga	gagtcattcc	3120
gtgtcagtat	tttacacaac	tgtttttctg	taaaggggga	aaaagaattc	aaatcttctc	3180
tttcaagaat	gctgactgct	gccaactgcc	tctcccgtg	gcccctctct	gtatagacag	3240
gcatagctat	ggtgaggact	tggtcggtc	ttgtctttct	cctctctctg	cttctctacc	3300
ctttctctcg	tgccctccac	ttaccaggcc	ctgggaagct	acacaccagg	caacagtgc	3360
cagggcctcg	gcctgggctt	cgaccaatta	ctagagcaga	aacagcagca	gctgcagtgt	3420
tggtttgtgc	tggtgcactgt	attaggttgt	gttttcatca	cctttgggtt	ttgtgtgtt	3480
ttgatgaagt	cctggtagca	ttctagtgtt	tacattctgg	gtagatagag	tttattcaag	3540
gtctcaaggc	atatgaatgg	aagagctcct	ctttacagcc	attcgtgtag	catgcataac	3600
tgtctctctg	tattctctct	agtgtctttt	ttttgtgtg	tgaatctgat	gtctgtttat	3660
tcacctacaa	tggtggagtaa	tggtcataaa	catataaagt	acttatgcct	ttatctgcca	3720
aattgtattt	aacttttcag	cttttaatat	aactttttat	ataataatta	atttattttt	3780
aaaaaaattg	aataccagcc	tggttatagt	gcatatgcct	gtgttcctag	cactcaggag	3840
acaaaggcag	aagtgtgaga	acttcagact	catactcagc	tatatacaag	accccaaatt	3900
tgtgctagat	tctgcagtag	agccatgagt	gtcccatct	tagagggaga	tcgctcatcc	3960
ttgtgctgtt	ctttaagtct	taccttgcaa	cccactgtaa	gtacactctt	gctcacagtc	4020
ctttagaatc	tcacactctt	tctctttaca	gacaccatgt	cattgcccac	tttattattt	4080
atctgatgtc	tacaaagatt	atgaaagaga	aacttgtag	cattctgtgt	aaagtacttg	4140
acacaaataa	tagtattcaa	gaatgacttc	ttaaatgaac	actgaatgaa	tagttgtgtc	4200
taattttttt	gatcaacaaa	tcaaaaaata	tttagattaa	atatctaaga	tacaaagcat	4260
aataccacat	gaatcattaa	agtgtgtaat	caatcttata	agtgtactgac	cctaaaactc	4320
atagacatta	ataattgctt	tcattgctta	gatataaact	ttattgatta	atacgttctc	4380
atgaaagtgg	ttcttggaag	gttctggaaa	cgaaaatatt	tttcttactg	cttttttctt	4440
ctagtaactg	attgaatttt	tctgcagttc	cataaagcat	ctgggtcaatt	gctattatcc	4500
aatatgagga	tatataacaa	agtattgatt	tttaaatgtg	gcgggtgata	gacaagactg	4560
ggcgtgtgaa	tgaggggggc	tctgtttctt	gtcccttctc	ttgggttctt	ttccttttgt	4620
tggttttgcct	tctccagctg	ctatgtgatg	ggttctgatt	atcttattat	atcttatttt	4680
gttatttttc	attgttatct	cttagaagcc	aacatgttat	atcacctcca	ctcccacat	4740
taggtgttct	cacaaatacc	ccaagctaaa	caaccacatc	atgtcatgtt	nctgtatact	4800
tccataagtg	ttgttaactc	tactgactct	tgtagcagg	cccaattggc	tttatcccta	4860
gctgggtgac	ctgggttctc	ccccaacacc	ataccgtcca	tcaaaactgag	tccttttcca	4920
agcacacacc	agatactgct	catctgagga	ctcttctcat	ccacctaaag	actgcctgct	4980
cctcggcaga	aagggcctct	agtcctcatc	ccttacgccc	tcaccaatgc	cttaggaaca	5040

tgtgtcfaat	gcccctgtgg	gtcatttccg	tttacagtag	ggaaatttgc	ctgataactt	5100
gcagcacacc	tataaagagg	ccttgcttgc	tctcatattt	agctggagaa	gataatgtac	5160
tcaccaactc	cactctatgc	aaccagctct	gctctgccc	tgccagtcag	acgtgaatct	5220
tacacctgga	ttcagattga	tgaatctaca	acatcaccca	ctccatgctt	ccttctaaat	5280
cagcagttct	agcctgaatg	acagatgcta	cccaagtctc	atctagttag	cctgtgccgg	5340
agtaaccctg	accttgagga	ttagaccagg	atgcacatcc	tgccaccagt	ccctttgtcc	5400
acctgacttc	atcccacccg	ggccatagcc	catgctcagg	ctccaccctc	catgcacaaa	5460
gctgggctttt	ccagcttcct	tcacctgtat	cagacacaaa	tagcaaaagg	ggccacagtg	5520
cctaggtccc	atcacaagac	catgtgcggt	agtttgaaa	acagtctcca	cttgaggctc	5580
agatagnttg	gaatcttggc	tctcatgtag	ttgtactgat	tagatcagtt	taggaagtac	5640
gacctttntg	gaaagaacat	ataactggga	ggggctttga	gatgtaaagg	ccccacacaa	5700
ttcctagtct	aaactntact	gcctgtctca	agcttgaggc	atgaactctc	actgttcctg	5760
atgtcattgt	tctgtctgct	ttccacaatt	ccctatcatt	aggggtccct	ttccttccctg	5820
gaatttttaag	tataaataaa	cncttctttn	taaaacaaca	agaacaacaa	acttgacnct	5880
gataatggat	tttaaggcgt	cttctctgga	taagaaaaaa	aaaagaatat	atgtgcatag	5940
gtgctgtatt	acttttgtca	ttgggtataac	ctgactggaa	gcaacttaaa	ggaaagaagaa	6000
tgtatcttga	ttttagattg	aagagcacca	tgactaagaa	ggcatagcag	cacaggtgca	6060
ccagcaagaa	cataggctgc	tagctcagat	ctctgtagat	atgggaacag	ggcagggaagc	6120
tagtagtcta	taaaccctcag	gacccatccc	atggagttcc	ttgtcttcca	gtgatgtcct	6180
gtgtcttaaa	gtttcacagt	tcccacagca	gcacctgccg	tctgggaacc	aacctgtggt	6240
ggatatttta	caacgtgata	ggcatatttt	gtctctagcc	ctgtaggttc	atagccatcc	6300
tatacttcag	tttatctagt	ccacctcagt	ctgatgggtc	tatagttcca	acacttcaaa	6360
actacaaagt	cttaaggggc	atgggctcgg	gtttattaga	gcagtaacac	ctctactagc	6420
tttctgtggt	acccactcct	cttaaggtct	ggttgaaatc	ctaataaggaa	gcagcttgag	6480
aggagggttt	attgtggccc	atactttgtt	ggtacattct	atcatgcaag	gggtggcactg	6540
tgatacagcc	gaggccatcc	gaggatggta	ctgttggtct	acatctgggt	gggacaggaa	6600
atggtaattc	tcaaggccca	cctgcttggt	gacttctttc	agttaagccc	catactctaa	6660
atcctctaca	acctcccaac	ataatgccac	cagctgggga	tcagctggtg	acagtgtctgg	6720
ccagggggag	cagttttaa	ccagaccagg	ggacctgaaa	acagagaact	gcagaggggc	6780
tgtgggactt	tataccagct	ttgcagacaa	atcacggcat	ttctttgtga	gcttgggtcca	6840
taaacaaaata	tatatctctc	tataggctcc	tttagtggtg	gtttcatatc	cacaaatttg	6900
ttcagaaaaa	cactgtgttt	tatgctagct	gtgtaggaga	taataccgct	gggagtcact	6960
tgagcatgga	taagtacat	agttcgtcct	catgagctcc	tgctctggtt	ctgtattatg	7020
tttacttgat	gagtttagtt	gtgcagttgg	ccaccaatta	aaaagtatca	ttttattttt	7080
tttacaatac	tcagttctca	agttaggagt	tttgttatta	tatggcttca	atattcacat	7140
tttaaccttt	ccaggagtta	agtataaaaa	cttatatcaa	ctgttgactt	agtaaatatc	7200
tattacagat	actataattc	tcttagttta	tatcatgaat	atgaggttgc	ttaaagtaag	7260
tgatgtaaaa	tacactaggg	gatgcttata	aaatggaatg	ttgtgagttt	tttgaaacac	7320
gagtactaaa	ttcataagtt	tttaaatagt	tacactgtta	gcttcagtag	tgctagatag	7380
atgtctataa	tggtctgaaga	gtggagcttg	gataattata	gtgtactctg	tatatctcat	7440
cagacatata	gcagattcca	ctagtatgtg	tggttaatat	gtgctaataa	aaatttaata	7500
caaaaagtc	gttttattac	tggaaccag	aggggttggt	tggtgctgatt	ttaatgcagt	7560
gactattagc	atattctaa	aaacagtttt	naggatttta	aagattggct	ttaccataaa	7620
tgtagagcta	tgttttacta	taatccatat	tatggctggc	cttaattcaa	tctctgcagt	7680
ttgggttact	tgctcaaagt	gaaggtcatt	tataaatgat	acacattttc	tcaccatagg	7740
aaatactacc	tggtccataa	cagagttaga	attgctaaat	tgatggtagc	aacaatggac	7800
tcaacacaaa	ctaaagttta	tttatgcca	cagatgtatc	ttgtgatttt	cccagaggga	7860
acaaggtata	atgcaacata	cacaaaactc	ctttcagcca	gtcaggcatt	tgctgctcag	7920
cggggtaagt	aaagatttaa	ctgtattcag	aaaaaacact	ttttaagaag	agtgtctttt	7980
gtttccttca	gagtcatact	aaagaatatg	cgtttcttgt	aagagctaa	tgagagaata	8040
tccgatcttc	tacagagtta	ggtatattct	tattagtctg	tgctctgagag	gttagagacg	8100
caggcttgct	atggcacatt	tcccatgctg	tgaattgagt	taaaaatgta	ggtaaatgat	8160
atccccaaga	aagtatactt	ttnggagtga	ctcagtataa	agcctgggtg	tataacataa	8220
acacgcacgt	gcggatgtat	atgtagcaca	tatgtaaaca	caggatatatg	canttgtaat	8280
aagaaagtgg	aggtcggggc	ccactgcagg	caagtctttt	agtgtgctg	agctaattgct	8340
gagaggtaga	aagaccaaga	aggctggagt	tgctcattcg	gcaaaggctg	gagctcactg	8400
tggtccataa	ctcagagtgt	ctgtctccct	tttgatagag	ttttcttgtt	tttaattatt	8460
agttttttaca	attatcccat	aaaatgtggg	ctcattgtgg	tcatcgtttt	cataaagtcc	8520
ttcaagtata	caccacagca	gtatctaaat	acactgggaa	gaatcagtc	gctgatggct	8580
tgaagtttca	ggacatctag	tgccacatca	tgcttcagaa	ccgacctgca	cttagtcagg	8640
gtcatattca	tgccacgtga	agacagagag	aggccatgcc	gtctgactta	ggatggaaat	8700

ttccttcgag	caaacacgaa	cgggctaggt	cttagttata	ggcatagtgt	ctgtgggttat	8760
actaggcaga	cattagtggg	ctgggtgtta	gaaggtacag	acaggcaaga	atttgctgta	8820
gatttgtttc	cctcatgtgt	tgacaccaca	tctaacctgc	tttttgagct	tctagtccta	8880
ataatctcat	aaaaatactg	gttgaaccag	aaatgggtgt	gcaaagctat	gatcccagct	8940
cctgggatct	aggggtgggag	gatcataaat	ttgaggccag	cttgggtctg	tcttagagaa	9000
aaaagaaaaa	taaaaagtct	ggtcaaggta	acatggagcc	tggaagtttc	acagggtgat	9060
tctgtaaaag	tcctgagaca	agatggcctc	tagtggcgaa	tgacttagct	gacaangaaa	9120
actttccag	cttgggtgac	ttttcagact	tcatacaagt	ttgtgaataa	attacactcc	9180
ttctgccctt	gggactgaac	tcagatatgt	ggttggtgga	atggctttct	ttcccacacc	9240
accctgcatt	ttaaaaattc	ttctgtagac	agtcccacca	tcctgtagct	gttcttcctt	9300
atgtcgccac	tttccctgga	gagaggcagt	gcagacttca	acccgcttct	ccctagtcgc	9360
tggtcatagc	acatcgaaag	acctagtgtc	tcctgtgaaa	ttgtaagtac	atcctggagt	9420
ccaggagagg	aggaagccga	acagagtggg	gggaatgctg	agtctctgtc	taagaaagac	9480
tgctgtctta	gcaagatgct	gctgctctcc	tgctgtgtct	ttcttgctcag	aacttatcaa	9540
agagaaggct	cgcagtgggt	cataatcttc	ccaaggacca	gccttcccag	cttctcgcag	9600
catatctcat	tcagttagat	gtttaatgga	tatgtgtcaa	tggggttgac	ctaagtgaga	9660
tggaatgta	tgtgagcatt	ctagggtgta	ggttatggca	ttaaacttta	atttccgtct	9720
atttgtggta	gttgataagt	aatttagatg	ttgactttca	tgtattccta	attatgacca	9780
cattgaatct	acctgctttc	taggccttgc	agtattaaaa	cacgtactga	caccaagaat	9840
aaaggccact	cacgttgctt	ttgattctat	gaagagtcat	ttagatgcaa	tttatgtgt	9900
cacagtgggt	tatgaaggga	atgagaaagg	ttcaggaaaa	tactcaaate	caccatccat	9960
gactggtaag	tcctgtattc	catagaagct	gaatagtaca	tggtacaggt	aagataaaact	10020
cttgtttgtt	cgctttgtct	agcttggttc	agtttggttt	tcagttagagg	gttccactat	10080
gaagctctgg	ctggccggga	actcactatg	tagaccaagc	tgcccttggg	ctccactaca	10140
cccagcacca	atcacccact	cttatncttt	tatgcntttt	tgtttttgc	ttgagctttc	10200
tttataacat	gtttgggaag	gacattgtca	ttatttaca	gaagaaatat	ggtcttttcc	10260
caacatgcta	gaatttaaag	actcagaact	cttgcccttg	tcagtgcaca	agtgcagaat	10320
gctgtgaagt	gacgtggctt	tgagtgcaga	tagttcagg	aactatagcc	acagactcaa	10380
catttgaaca	tggaacagg	tgagaacgga	gtgatggaag	attctggccc	cttccagaga	10440
attcatttta	gagagagatg	agagttagtaa	ggaagagaga	agagagagac	gtggtatttt	10500
gctgcagact	aaagagatct	cttataatcg	cagtactaag	gaggaagaag	cagaagatga	10560
tgactacagg	gccaggctga	acaatctagt	aaaatcctaa	gtcaggaaag	cagggctgag	10620
gtgcagctca	gtaggagagt	tggtgtctgc	cctacacaag	gcctggattt	agctcccagt	10680
agcaacgaag	ggaggcgagg	gtggggcaaaa	tcgaacactt	actcttgagg	actcccttta	10740
tgaatattac	cacactccag	taaatactct	ccagagattt	cagatgagat	tctgcttctt	10800
ggttaaacag	agggcaagaa	tattatgtca	cactgaacat	gggatgggaag	acatgttctg	10860
aggaatgtct	gcactccagt	gtgatgaaga	cttgaagttt	agggacattt	tcctctccctg	10920
gccccactca	ccccatctgt	attgagtatt	cccctagtgc	tcactcttat	ttgtatgtta	10980
actttcagga	aggggaagca	gattgatatt	caaaccagc	cagttttctt	aaatactttg	11040
tgatgggat	tggtcttgac	agtaaatgag	gaaatgtaaa	atgtaaaaga	ttctaatttt	11100
taatatttta	aaggtgaggt	tttctgttag	tacgcagagt	gagaggtttc	ttactgatgt	11160
ctgcgtacct	agaggaagga	tggtactctc	tccaaggctt	gctgttagaa	gtcagtgcac	11220
tggtgttaac	aagagatatg	tgctaagtgc	gttttaattt	cagcttaata	ctgcaaatca	11280
taagtgcata	gctttattgt	tttaaatctt	tttagtctta	atgtttcatt	tttaccataa	11340
gttactttgt	ataatcacaa	attctaaact	agtaagacgt	gaaattttct	tcttctttgt	11400
tagagtttct	ctgcaaacag	tgcccaaaac	ttcatattca	ctttgatcgt	atagacagaa	11460
atgaagttcc	agaggaacaa	gaacacatga	aaaagtggct	tcatgagcgc	tttgagataa	11520
aagataggta	agtgtgaaga	gctccagcat	ttagaaagtg	cagttcaacc	aaattttact	11580
ctcagatcct	gcttgaaagg	agtcttttta	tcttcattat	ttagtaataa	ctaatacatc	11640
ctgcatagac	aagaccacat	atacttaaat	gtagcatggt	tcatgggtgcg	ttacccttgt	11700
ttacaatta	agtttaacat	cctacatcag	tttgctgtgt	gatttctgta	ccatgacaac	11760
tcaacacagc	gatgcgttta	ttccaaagtc	gatagcacag	caaaagtga	actaaagtct	11820
gtattgtttc	aagaatgctt	ttgtgaaact	cgggttaaat	cttattctat	cctttcgtgt	11880
tcacattgta	cattttcatg	agtcactata	aaaatcatga	catggtggcc	tacctgcagt	11940
gtttgctgga	cagtaggctg	ctgtgtgata	agagcctttc	ctcttcagct	acacggggga	12000
cacgaggctt	tggtgttcaa	gactgaagca	cgggtgagca	caacaccttt	gtgtgtgtgg	12060
aagggaaggga	attgttcttt	tcataatgaa	attgtccctt	ttcttgagtt	agtagaaagt	12120
attacaagga	tagagagttg	aaatgaagct	ttatattaga	tttatgcctt	gtgtgtgtcac	12180
gtgtttctac	ctgacataac	ttttcaaccc	agccgctcag	gattattttg	atgatgggaa	12240
caatgtaaga	aggcctatgt	atcggttaact	cactgttgta	gctctgtgga	ancggnctnn	12300
caggcagtag	ggacgcttct	gtgcttttgt	gcctgtcctg	ctgttagaat	cttacagagg	12360

aggatgaatg aatgaccctt tttattttctc ttgtctgctt ttctaatttt atgggaataa 12420
gaacttttgg taggtctctg tcactggcct cttgttggtg agagacacct tgagcaaaagc 12480
aactcttctg agagaaagca tttagttggg gaattcctta cagcttcaga ggttgagttc 12540
gttttcatca tgctgaggac agggaggcac tcaggcagga gaagtagttg agagccacat 12600
tctgatctac aggcagagag agacagactg agcctggcat gggttcttgg aacctcaaaag 12660
cctctcatcc ctaccccatc tcccagcccc tatacacact tcctccaaca aggtacacc 12720
ttctaactct tcttaaagag tcaccacatc cagcgactaa gcattcagat atgtgaacct 12780
gtcggagcct ttcttactca gatcacctca ggaggaaaac tcctatgcta taagaatttc 12840
tttttcttgc catctttgaa agcttggttt ttgttgatta gatcctggcc tcacacatgc 12900
tcggcaatca ttttactgtt gagctccagc ctcagccgtt ttcattgggt tatgggagtc 12960
gagccatggg agagaagcta gaaggccttt cgttttatga gtcgggttgg tggaaacct 13020
tacagatgga agatttacia acaaaaatga agctggggcc atcaaggctc agcactcgct 13080
gctcttccag agagttcagg ttcatttctc agtaaccaca tgggtggctt gtaaatgtaa 13140
cttcatattc aatgaccctg acaccctctt ctggcctctg tgggcaccag acacaatcat 13200
gggtatcaga cacacacact agccaacacc catctacata aaagtatata anacatatct 13260
ttatcttaaa aatccccgaa gtcctcatat aatatcttag atccccgcg tgttttgatt 13320
tttgtttccc acgtggtgag gatataatat catgtccaaa ctgtaaggag tgaatgccct 13380
cccgtgcctc tcggacacct ctgcactcat ccaagttttc taaggagctg tacttgctca 13440
gcaagtactc aatacctaata aatggttta tgtttgtttc aacacaaaa atgtccaaaa 13500
ctgaaagatc aattctgttg ttttcttctt ggccataggt tgctcataga gttctatgat 13560
tcaccagatc cagaaagaag aaacaaatct cctgggaaaa gtgttcattc cagactaagt 13620
gtgaagaaga ctttaccctc agtggtgatc ttggggagtt tgactgcggt cactgctgatg 13680
acggagtcgg gaaggaaact gtacatgggc acctggttgt atggaacctc cctgggctgc 13740
ctgtggtttg ttattaaagc ataagcaagt agcaggctgc agtcacagtc tcttattgat 13800
ggctacacat tgtatcacat tgtttcttga attaaataag gagttttctt gttgttggtt 13860
tttttgtttt gttttgttct gttttaagcc ttgatgattg aacactggat aaagtagagt 13920
ttgtgaccac agccaacatg catttgattt ggggcaaaaca catgtggctt ttcagggtgc 13980
gggggttgcg gagacatgga agctaagtg agtttatgct gntttttttt ttttttnaa 14040
tgttttcatg aattaatgtc cacttgtaaa gattattgga tactttctgt aattcagaag 14100
gttgtatttt aacactagtt tgcagtatgt ttcgctatat tgggtatctt ccatttgact 14160
acttggcagc tcagactctt aatactaaag tattttacat tttgaagcta tgtgatactg 14220
gttttttgtt gttgttggg ttgttaattt ctgaaagtca atgaaagaca ctgtaatgat 14280
gcgttaagat gttccaagaa aaagggtgaga attattcatg gcaaaaaaga tctgtctagt 14340
gtatattttt atttatattg tctatttagc taattttctt tatatttgca aaataatgaa 14400
catttttaat atttattaaa atgcttgatt tgcatacccc cgattctaca gagaataatg 14460
tgtaaatgtg cagaatagac ttgaagctct gctgtgactc agtctccttt gtcagagctt 14520
ctagtagccc agctactgag ctgcttgggt agtacctcca gcacctgagc cgtaagtac 14580
ttataaatgc aagggaccgg ttatcttcat atcggaatag acatgaacag agctctaagg 14640
cgatgaaagt ctgccagcat cctctctgtc ctgcacagtg ccttctgcct ggctccattt 14700
gcttttggac tgcgttcgat cttagagtga ggtgctcact gcttatttca gccttggctc 14760
tgtggttttg tgcctccag ttgtgctgtt cactgttggg gtgcaggtgg tgctgccctg 14820
actcagaggg gcagctccct ggctcctgag ggtgagcctt cttggctact acagaagtat 14880
tgtgctgttg tgtatggcaa gaaccatcag gattggataa atgtgttatt tctcttggat 14940
ttccatggag ccacactgtt ggtacatgtc cctgtgaaac agagctacct ttcaggagca 15000
catcactctg tctgagtgca cggcacggtg tgcctgtga gaagaggctt tctaactgtt 15060
gatttgccgt gtttctatgt tgtgatttaa gcgtgattgc ctactagtca ttcaaggtaa 15120
aatatagacca attgaatcac aataaataat tttttttaat tgagggaaaa tttgcttctt 15180
gttttttcaa agccagaaaa cgagccattt caaacatctt tgaagagtca tgtgctgtca 15240
cttgttttct atgtgttagt gtctatattc atgtatggat acacatgaac atgtatatc 15300
atacacacac gccaatagaa tataacagcc taaaaacaat ccagcttggg tatcatgtta 15360
ctgtgctgaa ttgtaatggt ttttacttac aaagtgaggc taaaatcgat ttcattgctt 15420
tgttaaatac gtttttttca gcaatcctat tagagcttat ttgaccaga tcaaaaataag 15480
tacaagttca gagactttaa atatggctga ggtctagagc gatagctcag tagttaggaa 15540
cacatgccac tctttcaagg gcttcagttc ccagcactca tatggaggct cacagaaggc 15600
tgggaattcca gcttcatgga attggacaca tctcttagct tccatggatc tgtctgtctg 15660
tctctccctt ctctctctct ctctctctct ctctctctct ctctct 15720
15766

<210> 74

<211> 354

<212> PRT

<213> Mus musculus

<400> 74

```

Met Arg Tyr Leu Leu Pro Ser Val Leu Leu Leu Gly Ser Ala Pro Thr
1      5      10      15
Tyr Leu Leu Ala Trp Thr Leu Trp Arg Val Leu Ser Ala Leu Met Pro
      20      25      30
Ala Arg Leu Tyr Gln Arg Val Asp Asp Arg Leu Tyr Cys Val Tyr Gln
      35      40      45
Asn Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln Ile Leu
      50      55      60
Leu Tyr Gly Asp Leu Pro Lys Asn Lys Glu Asn Val Ile Tyr Leu Ala
      65      70      75      80
Asn His Gln Ser Thr Val Asp Trp Ile Val Ala Asp Met Leu Ala Ala
      85      90      95
Arg Gln Asp Ala Leu Gly His Val Arg Tyr Val Leu Lys Asp Lys Leu
      100      105      110
Lys Trp Leu Pro Leu Tyr Gly Phe Tyr Phe Ala Gln His Gly Gly Ile
      115      120      125
Tyr Val Lys Arg Ser Ala Lys Phe Asn Asp Lys Glu Met Arg Ser Lys
      130      135      140
Leu Gln Ser Tyr Val Asn Ala Gly Thr Pro Met Tyr Leu Val Ile Phe
      145      150      155      160
Pro Glu Gly Thr Arg Tyr Asn Ala Thr Tyr Thr Lys Leu Leu Ser Ala
      165      170      175
Ser Gln Ala Phe Ala Ala Gln Arg Gly Leu Ala Val Leu Lys His Val
      180      185      190
Leu Thr Pro Arg Ile Lys Ala Thr His Val Ala Phe Asp Ser Met Lys
      195      200      205
Ser His Leu Asp Ala Ile Tyr Asp Val Thr Val Val Tyr Glu Gly Asn
      210      215      220
Glu Lys Gly Ser Gly Lys Tyr Ser Asn Pro Pro Ser Met Thr Glu Phe
      225      230      235      240
Leu Cys Lys Gln Cys Pro Lys Leu His Ile His Phe Asp Arg Ile Asp
      245      250      255
Arg Asn Glu Val Pro Glu Glu Gln Glu His Met Lys Lys Trp Leu His
      260      265      270
Glu Arg Phe Glu Ile Lys Asp Arg Leu Leu Ile Glu Phe Tyr Asp Ser
      275      280      285
Pro Asp Pro Glu Arg Arg Asn Lys Phe Pro Gly Lys Ser Val His Ser
      290      295      300
Arg Leu Ser Val Lys Lys Thr Leu Pro Ser Val Leu Ile Leu Gly Ser
      305      310      315      320
Leu Thr Ala Val Met Leu Met Thr Glu Ser Gly Arg Lys Leu Tyr Met
      325      330      335
Gly Thr Trp Leu Tyr Gly Thr Leu Leu Gly Cys Leu Trp Phe Val Ile
      340      345      350

```

Lys Ala

<210> 75

<211> 22

<212> DNA

<213> Mus Musculus

<220>

<221> misc_binding

<222> 1..22

<223> amplification oligonucleotide g34292.pu

<400> 75

attaaaacac gtactgacac ca

<210> 76

<211> 19

<212> DNA

<213> Mus Musculus

<220>

```

<221> misc_binding
<222> 1..19
<223> amplification oligonucleotide g34292.rp
<400> 76
agtcattggat ggtggattt
<210> 77
<211> 26
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..26
<223> amplification oligonucleotide BOXIed
<400> 77
aatcatcaaa gcacagttga ctggat
<210> 78
<211> 33
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..33
<223> amplification oligonucleotide BOXIIer
<400> 78
ataaaccacc gtaacatcat aaattgcac taa
<210> 79
<211> 22
<212> DNA
<213> Mus Musculus
<220>
<221> misc_binding
<222> 1..22
<223> sequencing oligonucleotide moPGrace3S473
<400> 79
gagataaaaag ataggttgct ca
<210> 80
<211> 19
<212> DNA
<213> Mus Musculus
<220>
<221> misc_binding
<222> 1..19
<223> sequencing oligonucleotide moPGrace3S526
<400> 80
aagaaacaaa tttcctggg
<210> 81
<211> 18
<212> DNA
<213> Mus Musculus
<220>
<221> misc_binding
<222> 1..18
<223> sequencing oligonucleotide moPGrace3S597
<400> 81
tcttggggag tttgactg
<210> 82
<211> 18
<212> DNA
<213> Mus Musculus
<220>
<221> misc_binding

```

```

<222> 1..18
<223> sequencing oligonucleotide moPGrace5R323
<400> 82
gaccccggtg tagttctc
<210> 83
<211> 17
<212> DNA
<213> Mus Musculus
<220>
<221> misc_binding
<222> 1..17
<223> sequencing oligonucleotide moPGrace5R372
<400> 83
cagtaaagcc ggtcgtc
<210> 84
<211> 17
<212> DNA
<213> Mus Musculus
<220>
<221> misc_binding
<222> 1..17
<223> sequencing oligonucleotide moPGrace5R444
<400> 84
caggccagca ggtaggt
<210> 85
<211> 19
<212> DNA
<213> Mus Musculus
<220>
<221> misc_binding
<222> 1..19
<223> sequencing oligonucleotide moPGrace5R492
<400> 85
agcaggtagc gcatagagt
<210> 86
<211> 27
<212> DNA
<213> Mus Musculus
<220>
<221> misc_binding
<222> 1..27
<223> amplification oligonucleotide moPG13LR2
<400> 86
ggaaacaatg tgataaatg tgtagcc
<210> 87
<211> 18
<212> DNA
<213> Mus Musculus
<220>
<221> misc_binding
<222> 1..18
<223> amplification oligonucleotide moPG15
<400> 87
tggcgagccg agaggatg
<210> 88
<211> 36
<212> DNA
<213> Mus Musculus
<220>
<221> misc_binding
<222> 1..36

```

18

17

17

19

27

18

```

<223> amplification oligonucleotide moPG15BamI
<400> 88
cgtggatccg gaaacaatgt gatacaatgt gtagcc
<210> 89
<211> 27
<212> DNA
<213> Mus Musculus
<220>
<221> misc_binding
<222> 1..27
<223> amplification oligonucleotide moPG15EcoI
<400> 89
cgtgaattct ggcgagccga gaggatg
<210> 90
<211> 20
<212> DNA
<213> Mus Musculus
<220>
<221> misc_binding
<222> 1..20
<223> amplification oligonucleotide moPG1RACE3.18
<400> 90
ctgccagaca ggatgcccta
<210> 91
<211> 23
<212> DNA
<213> Mus Musculus
<220>
<221> misc_binding
<222> 1..23
<223> amplification oligonucleotide moPG1RACE3.63
<400> 91
acaagttaaa atggcttccg ctg
<210> 92
<211> 18
<212> DNA
<213> Mus Musculus
<220>
<221> misc_binding
<222> 1..18
<223> sequencing oligonucleotide moPG1RACE3R94
<400> 92
caaatgcatg ttggctgt
<210> 93
<211> 20
<212> DNA
<213> Mus Musculus
<220>
<221> misc_binding
<222> 1..20
<223> amplification oligonucleotide moPG1RACE5.276
<400> 93
gcaaatgcct gactggctga
<210> 94
<211> 22
<212> DNA
<213> Mus Musculus
<220>
<221> misc_binding
<222> 1..22
<223> amplification oligonucleotide moPG1RACE5.350

```

```

<400> 94
aatcaaaagc aacgtgagtg gc                22
<210> 95
<211> 20
<212> DNA
<213> Mus Musculus
<220>
<221> misc_binding
<222> 1..20
<223> amplification oligonucleotide moPG3RACE2
<400> 95
tgggcacctg gttgtatgga                20
<210> 96
<211> 20
<212> DNA
<213> Mus Musculus
<220>
<221> misc_binding
<222> 1..20
<223> amplification oligonucleotide moPG3RACE2n
<400> 96
tccttggtg cctgtggtt                20
<210> 97
<211> 21
<212> DNA
<213> Mus Musculus
<220>
<221> misc_binding
<222> 1..21
<223> sequencing oligonucleotide moPG3RACES20
<400> 97
gatggctaca cattgtatca c                21
<210> 98
<211> 24
<212> DNA
<213> Mus Musculus
<220>
<221> misc_binding
<222> 1..24
<223> sequencing oligonucleotide moPG3RACES5
<400> 98
tcctgaatta aataaggagt ttct                24
<210> 99
<211> 24
<212> DNA
<213> Mus Musculus
<220>
<221> misc_binding
<222> 1..24
<223> sequencing oligonucleotide moPG3RACES90
<400> 99
gtttgttatt aaagcataag caag                24
<210> 100
<211> 216
<212> DNA
<213> Homo sapiens
<400> 100
ctgctgtccc tgggtgtcca cacgtactcc atgcgctacc tgctgcccag cgtcgtgctc    60
ctgggcacgg cgcccaccta cgtgttggtc tggggggtct ggcggtgct ctccgccttc    120
ctgccgccc gcttctacca agcgctggac gaccggctgt actgcgtcta ccagagcatg    180
gtgctcttct tcttcgagaa ttacaccggg gtccag    216

```

```

<210> 101
<211> 70
<212> DNA
<213> Homo sapiens
<400> 101
atattgctat atggagattt gccaaaaaat aaagaaaata taatatattt agcaaatcat    60
caaagcacag                                         70
<210> 102
<211> 116
<212> DNA
<213> Homo sapiens
<400> 102
ttgactggat tgttctgac atcttggcca tcaggcagaa tgcgctagga catgtgctgct    60
acgtgctgaa agaaggggta aaatggctgc cattgtatgg gtgttacttt gctcag    116
<210> 103
<211> 90
<212> DNA
<213> Homo sapiens
<400> 103
catggaggaa tctatgtaaa gcgcagtgcc aaatttaacg agaaagagat gcgaaacaag    60
ttgcagagct acgtggacgc aggaactcca                                         90
<210> 104
<211> 91
<212> DNA
<213> Homo sapiens
<400> 104
atgtatcttg tgatttttcc agaagggtaca aggtataatc cagagcaaac aaaagtcctt    60
tcagctagtc aggcatttgc tgcccaacgt g                                         91
<210> 105
<211> 159
<212> DNA
<213> Homo sapiens
<400> 105
gccttgctgt attaaaacat gtgctaacac cacgaataaa ggcaactcac gttgcttttg    60
attgcatgaa gaattattta gatgcaattt atgatgttac ggtgggttat gaagggaaag    120
acgatggagg gcagcgaaga gagtcaccga ccatgacgg                                         159
<210> 106
<211> 124
<212> DNA
<213> Homo sapiens
<400> 106
aatttctctg caaagaatgt ccaaaaaattc atattcacat tgatcgtatc gacaaaaaag    60
atgtcccaga agaacaagaa catatgagaa gatggctgca tgaacgtttc gaaatcaaag    120
ataa                                         124
<210> 107
<211> 4342
<212> DNA
<213> Homo sapiens
<220>
<221> polyA_signal
<222> 325..330
<223> AATAAA potential
<221> polyA_signal
<222> 694..699
<223> AATAAA potential
<221> polyA_signal
<222> 828..833
<223> AATAAA potential
<221> polyA_signal
<222> 1821..1826
<223> AATAAA potential

```

```

<221> polyA_signal
<222> 2480..2485
<223> AATAAA : potential
<221> polyA_signal
<222> 2800..2805
<223> AATAAA : potential
<221> polyA_signal
<222> 4264..4269
<223> AATAAA : potential
<221> polyA_signal
<222> 4320..4315
<223> AATAAA
<400> 107
gatgcttata gaattttatg agtcaccaga tccagaaaga agaaaaagat ttcttgggaa 60
aagtgttaat tccaaattaa gtatcaagaa gactttacca tcaatgttga tcttaagtgg 120
tttgactgca ggcatgctta tgaccgatgc tggaaggag ctgtatgtga acacctggat 180
atatggaacc ctacttggct gcctgtgggt tactattaaa gcatagacaa gtacgtgtct 240
ccagacagtg ggatgtgcta cattgtctat ttttggcggc tgcacatgac atcaaattgt 300
ttcctgaatt tattaaggag tgtaaataaa gccttgttga ttgaagattg gataatagaa 360
tttgtgacga aagctgatat gcaatggctc tggggcaaca tacctgggtg tacaacttta 420
gcctcggggc tgctggaagg gtaaaagcta aatggagttt ctctgtctct gtccatttcc 480
tatgaactaa tgacaacttg agaaggctgg gaggattgtg tattttgcaa gtcagatggc 540
tgcatttttg agcattaatt tgcagcgtat ttacttttt ctgttatttt caatttatta 600
caacttgaca gctccaagct cttattacta aagtatttag tatcttgacg ctagttaata 660
tttcatcttt tgcttatttc tacaagtcag tgaataaaat tgtatttagg aagtgtcagg 720
atgttcaaag gaaagggtaa aaagtgttca tggggaaaaa gctctgttta gcacatgatt 780
ttattgtatt gcgttattag ctgattttac tcattttata tttgcaaaat aaatttctaa 840
tatttattga aattgcttaa tttgcacacc ctgtacacac agaaaatggg ataaaaatag 900
agaacgaagt ttaaaattgt gactctgatt cattatagca gaactttaaa tttcccagct 960
ttttgaagat ttaagctacg ctatttagac ttccctttgt ctgtgccata agtcttgaa 1020
aacgttaagg ttttctgttt tgttttgttt ttttaataac aaaagagtcg gtgtgaacct 1080
tggttgacc ccaagttcac aagattttta agtgatgag agcctgcaga cattctgcct 1140
agatttacta gcgtgtgect tttgcctgct tctctttgat ttcacagaat attcattcag 1200
aagtcgcgtt tctgtagtgt ggtggattcc cactgggctc tggctcctcc cttggatccc 1260
gtcagtggtg ctgctcagcg gcttgcacgt agacttgcta ggaagaaatg cagagccagc 1320
ctgtgctgcc cactttcaga ttggaactct ttaagccctt gtgagtgagg ttcaccagct 1380
actgcagagg cattttgcat ttgtctgtgt caagaagttc accttctcaa gccagtgaag 1440
tacagactta attcgtcatg actgaacgaa tttgtttatt tcccattagg tttagtggag 1500
ctacacatta atatgtatcg ccttagagca agagctgtgt tccaggaacc agatcacgat 1560
ttttagccat ggaacaatat atcccattgg agaagacctt tcagtgtgaa ctgttctatt 1620
tttgtgttat aatttaaaact tctgatttct catagtcctt taagttgaca tttctgttta 1680
ctgctactgg attttttctg cagaaatata tcagtggccc acattaaaca taccagttag 1740
atcatgataa gcaaaatgaa agaaataatg attaaggga aattaagtga ctgtgttaca 1800
ctgcttctcc catgccagag aataaaactct ttcaagcatc atctttgaag agtcgtgtgg 1860
tgtgaattgg tttgtgtaca ttagaatgta tgcacacatc catggacact caggatatag 1920
ttggcctaata aatcggggca tgggtaaaac ttatgaaaat ttctcatgca tgaattgtaa 1980
ttttctctta cctgtaaaag taaaatttag tcaattccat gtctttgtta agtacagga 2040
tttaatatat tttgaatata atgggtatgt tctaaatttg aactttgaga ggcaatactg 2100
ttggaattat gtggattcta actcatttta acaaggtagc ctgacctgca taagatcact 2160
tgaatgttag gtttcataga actatactaa tcttctcaca aaaggtctat aaaatacagt 2220
cggtgaaaaa aattttgtat caaaatgttt ggaaaattag aagcttctcc ttaacctgta 2280
ttgatactga cttgaattat tttctaaaat taagagccgt atacctacct gtaagtcttt 2340
tcacatatca tttaaacttt tgtttgtatt attactgatt tacagcttag ttattaattt 2400
ttctttataa gaatgccgtc gatgtgcatg cttttatggt tttcagaaaa ggggtgtgtt 2460
ggatgaaagt aaaaaaaaaa ataaaatctt tcactgtctc taatggctgt gctgtttaac 2520
attttttgac cctaaaattc accaacagtc tcccagtaca taaaatagga ttaatgactg 2580
gccctgcatt cttcacataa tttttcccta agctttgagc aaagttttaa aaaaatacac 2640
taaaataatc aaaactgtta agcagtatat tagtttgggt atataaaatc atctgcaatt 2700
tataagatgc atggccgatg ttaatttgct aagcaattct gtaatcatta agtgatctca 2760
gtgaaacatg tcaaatgcct taaatttaact aagttgggtg ataaaagtgc cgatctggct 2820
aactcttaca ccatacatat tgatagtttt tcatatgttt catttccatg tgatttttaa 2880

```



```

aatttagagt ggcaacaatt ttgcttaata tgggttacat aagctttatt ttttcctttg 2940
ttcataatta tttcttttga ataggtctgt gtcaatcaag tgatctaact agactgatca 3000
tagatagaag gaaataaggc caagttcaag accagcctgg gcaacataac gagaacctgt 3060
ctacaaaaaa attaaaaaaa attagccagg catggtggcg tacactgagt agtttgtccc 3120
agctactcgg gaggggtgagg tgggaggatc gcttcagccc aggaggttga gattgcagtg 3180
agccatggac ataccactgc actacagcct aggtaacagc acgagacccc aactcttaga 3240
aaatgaaaag gaaatataga aatataaaat ttgcttatta tagacacaca gtaactccca 3300
gatattgacc acaaaaaatg tgaaaagaga gagaatgtc taccaaagca gtattttgtg 3360
tgtataattg caagcgcata gtaaaataat ttaacctta atttggtttt agtagtggtt 3420
agattgaaga ttgagtgaat tttttcttg gcagatattc cgtatctggg ggaaagctac 3480
aatgcaatgt cgttgtagtt ttgcatggct tgctttataa acaagatttt ttctccctcc 3540
ttttgggcca gttttcatta cgagtaactc acactttttg attaaagaac ttgaaattac 3600
gttatcactt agtataattg acattatata gagactatgt aacatgcaat cattagaatc 3660
aaaattagta ctttgggtcaa aatatttaca acattcacat acttggtcaa tattcatgta 3720
attaactgaa tttaaaacct tcaactatta tgaagtgtc gtctgtacaa tcgctaattt 3780
actcagtta gagtagctac aactcttcca tactatcac aatatttgac atcttttcca 3840
atttgtgtat gaaaagtaaa tctattcctg tagcaactgg ggagtcatat atgaggtcaa 3900
agacatatat cttgttatta taatatgtat actataataa tagctgggta tcctgagcag 3960
gggaaaagggt tatttttagg aaaaccactt caaatagaaa gctgaagtac ttctaataa 4020
ctgagggag tataatatgt ggaacaaact ctcaacaaaa tgtttattga tgttgatgaa 4080
acagatcagt ttttccatcc ggattattat tggttcatga ttttatatgt gaatatgtaa 4140
gatattgtct gcaattttat aaatgttcat gtcttttttt aaaaaagggt ctattgaaat 4200
tctgtgtctc cagcaggcaa gaatacttga ctaactcttt ttgtctcttt atggtatttt 4260
cagaataaag tctgacttgt gtttttgaga ttattggtgc ctcattaatt cagcaataaa 4320
ggaaaatatg catctcaaaa at 4342
<210> 108
<211> 62
<212> DNA
<213> Homo sapiens
<400> 108
agattggatt cgtagattaa acttgagaaa caaaccataa aagtggagg ccctctttaa 60
ca 62
<210> 109
<211> 86
<212> DNA
<213> Homo sapiens
<400> 109
gagatggggg tctcgctgtg ttgcccaggc tgggtcttga ctcaagcaat ctgctgtct 60
cagcctacca aaatgctgga ttatag 86
<210> 110
<211> 116
<212> DNA
<213> Homo sapiens
<400> 110
gctaaagcag tctctctgag tagttaggac tacagacata cacgtgccac cgcgccagc 60
tccgtgttct ctttgtttcc ctgctctctg ctcttccact tatctttgca tggcag 116
<210> 111
<211> 45
<212> DNA
<213> Homo sapiens
<400> 111
ggaaagacga tggagggcag cgaagagagt caccgacat gacgg 45
<210> 112
<211> 5138
<212> DNA
<213> Homo sapiens
<220>
<221> misc_feature
<222> 31..33
<223> ATG
<221> misc_feature

```

<222> 262..264
 <223> TAG
 <221> polyA_signal
 <222> 5111..5116
 <223> AATAAA
 <400> 112

ctgctgtccc tgggtgetcca cacgtactcc atg cgc tac ctg ctg ccc agc gtc 54
 Met Arg Tyr Leu Leu Pro Ser Val
 1 5

gtg ctc ctg gcc acg gcg ccc acc tac gtg ttg gcc tgg ggg gtc tgg 102
 Val Leu Leu Gly Thr Ala Pro Thr Tyr Val Leu Ala Trp Gly Val Trp
 10 15 20

cgg ctg ctc tcc gcc ttc ctg ccc gcc cgc ttc tac caa gcg ctg gac 150
 Arg Leu Leu Ser Ala Phe Leu Pro Ala Arg Phe Tyr Gln Ala Leu Asp
 25 30 35 40

gac cgg ctg tac tgc gtc tac cag agc atg gtg ctc ttc ttc ttc gag 198
 Asp Arg Leu Tyr Cys Val Tyr Gln Ser Met Val Leu Phe Phe Phe Glu
 45 50 55

aat tac acc ggg gtc cag ttg act gga ttg ttg ctg aca tct tgg cca 246
 Asn Tyr Thr Gly Val Gln Leu Thr Gly Leu Leu Leu Thr Ser Trp Pro
 60 65 70

tca ggc aga atg cgc tag gacatgtgctg ctacgtgctg aaagaagggt 294
 Ser Gly Arg Met Arg *
 75

taaaatggct gccattgtat ggggtgttact ttgctcagca tggaggaatc tatgtaaagc 354
 gcagtgccaa atttaacgag aaagagatgc gaaacaagtt gcagagctac gtggacgcag 414
 gaactccaat gtatcttgtg atttttccag aaggtacaag gtataatcca gagcaaacaa 474
 aagtcctttc agctagtcag gcatttgcctg cccaacgtgg ccttgcagta ttaaaacatg 534
 tgctaacacc acgaataaag gcaactcacg ttgcttttga ttgcatgaag aattatttag 594
 atgcaattta tgatgttacg gtgggtttatg aagggaaaga cgatggaggg cagcgaagag 654
 agtcaccgac catgacggaa tttctctgca aagaatgtcc aaaaattcat attcacattg 714
 atcgtatcga caaaaaagat gtcccagaag aacaagaaca tatgagaaga tggctgcatg 774
 aacgtttcga aatcaaaagat aagatgctta tagaatttta tgagtacca gatccagaaa 834
 gaagaaaaag atttcctggg aaaagtgtta attccaaatt aagtatcaag aagactttac 894
 catcaatgtt gatcttaagt ggtttgactg caggcatgct tatgaccgat gctggaagga 954
 agctgtatgt gaacacctgg atatatggaa cctacttgg ctgcctgtgg gttactatta 1014
 aagcatagac aagtagctgt ctccagacag tgggatgtgc tacattgtct atttttggcg 1074
 gctgcacatg acatcaaatt gtttcttgaa tttatttaagg agtgtaaata aagccttgtt 1134
 gattgaagat tggataatag aatttgtgac gaaagctgat atgcaatggt cttgggcaaa 1194
 catacctggt tgtacaactt tagcatcggg gctgctggaa gggtaaaagc taaatggagt 1254
 ttctcctgct ctgtccattt cctatgaact aatgacaact tgagaaggct gggaggattg 1314
 tgtattttgc aagtcagatg gctgcatttt tgagcattaa tttgcagcgt atttcacttt 1374
 ttctgttatt ttcaatttat tacaacttga cagctccaag ctcttattac taaagtattt 1434
 agtatcctgc agctagttaa tatttcattt tttgcttatt tctacaagtc agtgaaataa 1494
 attgtattta ggpagtgtca ggatgttcaa aggaaagggt aaaaagtgtt catggggaaa 1554
 aagctctggt tagcacatga ttttattgta ttgcgttatt agctgatttt actcatttta 1614
 tatttgcaaa atpaattttt aatattttat gaaattgctt aatttgcaca ccctgtacac 1674
 acagaaaatg gtataaaata tgagaacgaa gtttaaaatt gtgactctga ttcattatag 1734
 cagaacttta aaktttccag ctttttgaag atttaagcta cgctatttagt acttcccttt 1794
 gtctgtgcca taagtgtctg aaaacgttaa ggttttctgt tttgttttgt ttttttaata 1854
 tcaaaagagt cgggtgtgaac cttggttgga cccaagtgc acaagatttt taagggtgatg 1914
 agagcctgca gacattctgc ctgattttac tagcgtgtgc cttttgctg cttctctttg 1974
 atttcacaga atattcattc agaagtcgag tttctgtagt gtggtggatt cccactgggc 2034
 tctggctcctt cccttggatc cgtcagtggt tgctgctcag cggttgacac gtgactgtgc 2094
 taggaagaaa tgcagagcca gcctgtgctg cccactttca gagggtgaact ctttaagccc 2154
 ttgtgagtgg gcttcaccag ctactgcaga ggcattttgc atttgtctgt gtcaagaagt 2214
 tcacctttctc aagccagtga aatacagact taattcgtca tgactgaacg aatttgttta 2274
 tttcccatga ggttttagtg agctacacat taatatgtat cgccttagag caagagctgt 2334
 gttccaggaa ccagatcacg attttttagcc atggaacaat atatcccatg ggagaagacc 2394
 tttcagtgtg aactgttcta ttttgtgtt ataatttaaa cttcgatttc ctcatagtcc 2454
 ttttaagttga catttctgct tactgtact ggatttttgc tgcagaaata tatcagtggc 2514

```

ccacattaaa cataccagtt ggatcatgat aagcaaaatg aaagaaataa tgattaaggg 2574
aaaattaagt gactgtgtta cactgcttct cccatgccag agaataaact ctttcaagca 2634
tcactcttga agagtcgtgt ggtgtgaatt ggtttgtgta cattagaatg tatgcacaca 2694
tccatggaca ctccaggatat agttggccta ataatcgggg catgggtaaa acttatgaaa 2754
atttcctcat gctgaattgt aattttctct tacctgtaaa gtaaaattta gatcaattcc 2814
atgtctttgt taagtacagg gatttaatat attttgaata taatgggtat gttctaaatt 2874
tgaactttga gaggcaatac tgttgggaatt atgtggattc taactcattt taacaaggta 2934
gcctgacctg cataagatca ctggaatgtt aggtttcata gaactatact aatcttctca 2994
caaaaggtct ataaaatata gtcgttgaaa aaaattttgt atcaaaatgt ttggaaaatt 3054
agaagcttct ccttaacctg tattgatact gacttgaatt attttctaaa attaagagcc 3114
gtatacctac ctgtaagtct tttcacatat catttaaact tttgtttgta ttattactga 3174
tttacagctt agttattaat ttttcttctt aagaatgccg tcgatgtgca tgcttttatg 3234
tttttcagaa aaggggtgtg ttggatgaaa gtaaaaaaaa aaataaaact tttcactgtc 3294
tctaattggc gtgctgttta acattttttg accctaaaaa tcaccaacag tctccagta 3354
cataaaatag gcttaatgac tggccctgca ttcttcacaa tattttctcc taagctttga 3414
gcaaagtttt aaaaaaatac actaaaataa tcaaaactgt taagcagtat attagtttgg 3474
ttatataaat tcacttgcaa ttataagat gcatggccga tgttaatttg cttggcaatt 3534
ctgtaatcat taagtgtctt cagtgaacaa tgtcaaatgc cttaaatata ctaagttggt 3594
gaataaaagt gccgatctgg ctaactctta caccatacat actgatagtt tttcatatgt 3654
ttcatttcca tgtgattttt aaaattttaga gtggcaacaa ttttgcttaa tatgggttac 3714
ataagcttta tttttctctt tgttcataat tatattcttt gaataggctt gtgtcaatca 3774
agtgtactaa ctgactgat catagataga aggaataaag gccaaagtca agaccagcct 3834
gggcaacata tcgagaacct gtctacaaaa aaattaaaaa aaattagcca ggcattgtgg 3894
cgtacactga gtagtttgtc ccagctactc gggagggtga ggtgggagga tcgcttcagc 3954
ccaggagggt gagattgcag tgagccatgg acataccact gcactacagc ctaggtaaca 4014
gcacgagacc ccaactctta gaaaatgaaa aggaatatata gaaatataaa atttgcttat 4074
tatagacaca cagtaactcc cagatatgta ccacaaaaaa tgtgaaaaga gagagaaatg 4134
tctaccaaag cagtattttg tgtgtataat tgcaagcgca tagtaaaata attttaacct 4194
taattttgtt ttagtatgtt tttagattgaa gattgagtga aatatattct tggcagatat 4254
tcggtatctg gtggaaagct acaatgcaat gtcgtttagt ttttgcatgg cttgctttat 4314
aaacaagatt ttttctccct ccttttgggc cagttttcat tacgagtaac tcacactttt 4374
tgattaaaga acttgaattt acgttatcac ttagtataat tgacattata tagagactat 4434
gtaacatgca atcattagaa tcaaaattag tactttggtc aaaatattta caacattcac 4494
atacttgtca aatattcatg taattaactg aattttaaac cttcaactat tatgaagtgc 4554
tcgtctgtac aatcgctaatt ttactcagtt tagagtagct acaactcttc gatactatca 4614
tcaatatttg acatcttttc caattttgtg atgaaaagta aatctattcc tgtagcaact 4674
ggggagtc atatgaggtc aaagacatat accttggtat tataatatgt atactataat 4734
aatagctggt tatcctgagc aggggaaaag gttattttta ggaataaccac ttcaaataga 4794
aagctgaagt acttctaata tactgaggga agtataatat gtggaacaaa cttcaacaa 4854
aatgtttatt gatgttgatg aaacagatca gtttttccat ccggtattat attggttcat 4914
gattttatat gtgaatatgt aagatatgtt ctgcaatttt ataaatgttc atgtcttttt 4974
ttaaaaaagg tgctattgaa attctgtgtc tccagcaggc aagaatactt gactaactct 5034
ttttgtctct ttatgggtatt ttcagaataa agtctgactt gtgtttttga gattattggt 5094
gcctcattaa ttcagcaata aaggaaaata tgcactctca aaat 5138

```

<210> 113

<211> 5224

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> 31..33

<223> ATG

<221> misc_feature

<222> 262..264

<223> TAG

<221> polyA_signal

<222> 5197..5202

<223> AATAAA

<400> 113

```

ctgctgtccc tgggtgtcca cacgtactcc atg cgc tac ctg ctg ccc agc gtc
Met Arg Tyr Leu Leu Pro Ser Val

```

54

gtg	ctc	ctg	ggc	acg	gcg	ccc	acc	tac	gtg	ttg	gcc	tgg	ggg	gtc	ttg	102
Val	Leu	Leu	Gly	Thr	Ala	Pro	Thr	Tyr	Val	Leu	Ala	Trp	Gly	Val	Trp	
10						15					20					
cgg	ctg	ctc	tcc	gcc	ttc	ctg	ccc	gcc	cgc	ttc	tac	caa	gcg	ctg	gac	150
Arg	Leu	Leu	Ser	Ala	Phe	Leu	Pro	Ala	Arg	Phe	Tyr	Gln	Ala	Leu	Asp	
25					30					35					40	
gac	cgg	ctg	tac	tgc	gtc	tac	cag	agc	atg	gtg	ctc	ttc	ttc	ttc	gag	198
Asp	Arg	Leu	Tyr	Cys	Val	Tyr	Gln	Ser	Met	Val	Leu	Phe	Phe	Phe	Glu	
				45					50					55		
aat	tac	acc	ggg	gtc	cag	ttg	act	gga	ttg	ttg	ctg	aca	tct	tgg	cca	246
Asn	Tyr	Thr	Gly	Val	Gln	Leu	Thr	Gly	Leu	Leu	Leu	Thr	Ser	Trp	Pro	
			60					65					70			
tca	ggc	aga	atg	cgc	tag	gacatgtg	cg	ctacgtgt	ctg	aaagaagg	gt					294
Ser	Gly	Arg	Met	Arg	*											
75																
taaaatggct	gccattgtat	gggtgttact	ttgctcagga	gatgggggtc	tcgctgtgtt	354										
gcccaggctg	gtcttggact	caagcaatct	gcctgtctca	gcctaccaaa	atgctgggatt	414										
atagcatgga	ggaatctatg	taaagcgcag	tgccaaattt	aacgagaaa	agatgcgaaa	474										
caagttgcag	agctacgttg	acgcaggaac	tccaatgtat	cttctgattt	ttccagaagg	534										
tacaagggtat	aatccagagc	aaacaaaaagt	cctttcagct	agtcaggcat	ttgctgccca	594										
acgtggcctt	cgagtattaa	aacattgtct	aacaccacga	ataaaggcaa	ctcactgttc	654										
ttttgattgc	atgaagaatt	atttagatgc	aatttatgat	gttcagggtg	tttatgaagg	714										
gaaaagcagt	ggagggcgagc	gaagagagtc	accgaccatg	acggaatttc	tctgcaaaag	774										
atgtccaaaa	attcatattc	acattgatcg	tatcgacaaa	aaagatgtcc	cagaagaaca	834										
agaacatatg	agaagatggc	tgcattgaag	tttcgaaatc	aaagatatac	gtccttataga	894										
attttatgag	tcaccagatc	cagaaagaag	aaaaagattt	cctgggaaaa	gtgttaattc	954										
caaattaagt	atcaagaaga	ctttaccatc	aatgttgatc	ttaaagtggt	tgactgcagg	1014										
catgcttatg	accgatgttg	gaaggaaagc	gtatgtgaac	acctggatat	atggaaccct	1074										
acttggtctg	ctgtggctga	ctattaaagc	atagacaagc	agctgtctcc	agacagtggt	1134										
atgtgtctaca	tgtgtcattt	tgtggcgctg	cacatgacat	aactattgtt	cctgaattta	1194										
ttaaggagtg	taataaaagc	ctgtgtgatt	gaagattgga	taatagaatt	tgtagcgaag	1254										
gctgatatgc	aatgggtctg	ggcacaacata	cctgggtgtg	caacttttag	atcggggctg	1314										
ctggaagggt	aaaagctaaa	tggagtttct	ctgctctgtg	ccatttctca	tgaataatg	1374										
acaacttgag	aaggctggga	gatttgtgtg	ttttgcaagt	cagatggctg	catttttgag	1434										
cattaatttg	cagcgtattt	cactttttct	gttattttca	atttattaca	acttgacagc	1494										
tccaagctct	tattactaaa	gtatttagta	tcttgacagt	agttaaatat	tcattctttg	1554										
ctttatttcta	caagtcagtg	aaataaaatt	tatttaggaa	gtgtcaggat	gttcaaaagg	1614										
aagggtaaaa	agtgttcatt	gggaaaaaagc	tctgttttag	acatgatttt	attgtattgc	1674										

```

ggattctaac tcattttaac aaggtagcct gacctgcata agatcacttg aatgttaggt 3054
ttcatagaac tataactaatc ttctcacaac aggtctataa aatacagtcg ttgaaaaaaa 3114
ttttgtatca aaatgtttgg aaaattagaa gcttctcctt aacctgtatt gatactgact 3174
tgaattattt tctaaaatta agagccgtat acctacctgt aagtcctttc acatatcatt 3234
taaaactttg ttgtattat tactgattta cagcttagtt attaatTTTT cttataaga 3294
atgccgtcga tgtgcatgct tttatgtttt tcagaaaagg gtgtgtttgg atgaaagtaa 3354
aaaaaaaaat aaaatctttc actgtctcta atggctgtgc tgtttaacat tttttgaccc 3414
taaaattcac caacagtctc ccagtacata aaataggcct aatgactggc cctgcattct 3474
tcacaattatt ttccctaag ctttgagcaa agttttaaaa aaatacacta aaataatcaa 3534
aactgttaag cagtatatta gtttggttat ataaattcat ctgcaattta taagatgcat 3594
ggccgatgtt aatttgcttg gcaattctgt aatcattaag tgatctcagt gaaacatgtc 3654
aaatgcctta aattaactaa gttggtgaat aaaagtgcg atctggctaa ctcttacacc 3714
atacactctg atagtttttc atagtgttca tttccatgtg atttttaaaa tttagagtgg 3774
caacaatttt gcttaatatg ggttacataa gctttatttt ttcccttggt cataattata 3834
ttctttgaat aggtctgtgt caatcaagtg atctaactag actgacata gatagaagga 3894
aataaggcca agttcaagac cagcctgggc aacatatcga gaacctgtct acaaaaaaat 3954
taaaaaaaat tagccaggca tgggtggcgt cactgagtag tttgtcccag ctactcggga 4014
gggtgaggtg ggaggatcgc ttccagcccag gaggttgaga ttgcagtgag ccatggacat 4074
accactgcac tacagcctag gtaacagcac gagaccccaa ctcttagaaa atgaaaagga 4134
aatatagaaa tataaaattt gcttattata gacacacagt aactcccaga tatgtaccac 4194
aaaaaatgtg aaaagagaga gaaatgtcta ccaaagcagt attttgtgtg tataattgca 4254
agcgcatagt aaaataattt taaccttaat ttgtttttag tagtgtttag attgaagatt 4314
gagtgaataa tttctctggc agatattccg tatctggttg aaagctacaa tgcaatgtcg 4374
ttgtagtttt gcatggcttg ctttataaac aagatttttt ctccctcctt ttgggccagt 4434
tttcattacg agtaactcac actttttgat taaagaactt gaaattacgt tatcacttag 4494
tataattgac atatataga gactatgtaa catgcaatca ttagaatcaa aattagtagt 4554
ttggtcaaaa tatttacaac attcacatac ttgtcaataa ttcatgtaat taactgaatt 4614
taaaaccttc aactattatg aagtgtctgt ctgtacaatc gctaatttac tcagtttaga 4674
gtagctacaa ctcttcgata ctatcatcaa tatttgacat cttttccaat ttgtgtatga 4734
aaagtataatc tttctctgta gcaactgggg agtcatatat gaggtcaaaag acatatacct 4794
tggtattata atagtatac tataataata gctggttatc ctgagcaggg gaaaagggtta 4854
tttttaggaa aaccacttca aatagaaagc tgaagtactt ctaatatact gaggggaagta 4914
taatatgtgg aacaaactct caacaaaatg tttattgatg ttgatgaaac agatcagttt 4974
ttccatccgg attattattg gttcatgatt ttatagtga atagtgaaga tatgttctgc 5034
aattttataa atgttcatgt ctttttttaa aaaagtgct attgaaattc tgtgtctcca 5094
gcaggcaaga atacttgact aactcttttt gtctctttat ggtattttca gaataaagtc 5154
tgacttgtgt ttttgagatt attggtgcct cattaattca gcaataaagg aaaatatgca 5214
tctcaaaaat 5224

<210> 114
<211> 4863
<212> DNA
<213> Homo sapiens
<220>
<221> misc_feature
<222> 31..33
<223> ATG
<221> misc_feature
<222> 745..747
<223> TAG
<221> polyA_signal
<222> 4836..4841
<223> AATAAA
<400> 114
ctgctgtccc tgggtgctcca cacgtactcc atg cgc tac ctg ctg ccc agc gtc 54
Met Arg Tyr Leu Leu Pro Ser Val
1 5
gtg ctc ctg ggc acg gcg ccc acc tac gtg ttg gcc tgg ggg gtc tgg 102
Val Leu Leu Gly Thr Ala Pro Thr Tyr Val Leu Ala Trp Gly Val Trp
10 15 20
cgg ctg ctc tcc gcc ttc ctg ccc gcc cgc ttc tac caa gcg ctg gac 150
Arg Leu Leu Ser Ala Phe Leu Pro Ala Arg Phe Tyr Gln Ala Leu Asp

```


atcgcccttag	agcaagagct	gtgttccagg	aaccagatca	cgatttttag	ccatggaaca	2097
atatatccca	tgggagaaga	cctttcagtg	tgaactgttc	tatttttgtg	ttataattta	2157
aacttcgatt	tcctcatagt	cctttaagtt	gacatttctg	cttactgcta	ctggattttt	2217
gctgcagaaa	tatatcagtg	gcccacatta	aacataccag	ttggatcatg	ataagcaaaa	2277
tgaaagaaat	aatgattaag	ggaaaattaa	gtgactgtgt	tacactgctt	ctcccatgcc	2337
agagaataaa	ctctttcaag	catcatcttt	gaagagtcgt	gtggtgtgaa	ttggtttgtg	2397
tacattagaa	tgtaggcaca	catccatgga	cactcaggat	atagttggcc	taataatcgg	2457
ggcatgggta	aaacttatga	aaatttcctc	atgctgaatt	gtaattttct	cttacctgta	2517
aagtaaaaat	tagatcaatt	ccatgtcttt	gttaagtaca	gggatttaat	atattttgaa	2577
tataatgggt	atgttctaaa	tttgaacttt	gagaggcaat	actgttgga	ttatgtggat	2637
tctaactcat	tttaacaagg	tagcctgacc	tgcatagaat	cacttgaatg	ttaggtttca	2697
tagaactata	ctaactctct	cacaaaagg	ctataaaaata	cagtcgttga	aaaaaatttt	2757
gtatacaaat	gtttggaaaa	ttagaagctt	ctccttaacc	tgtattgata	tgacttgaa	2817
ttattttcta	aaattaagag	ccgtatacct	acctgtaagt	cttttcacat	atcatttaaa	2877
cttttctgtg	tattattact	gatttacagc	ttagttatta	atttttcttt	ataagaatgc	2937
cgctgatgtg	catgctttta	tggttttcag	aaaagggtgt	gtttggatga	aagtaaaaaa	2997
aaaaataaaa	tccttctactg	tctctaattg	ctgtgctgtt	taacattttt	tgaccctaaa	3057
attcaccaac	agtctcccag	tacataaaaat	aggcttaatg	actggccctg	cattcttcac	3117
aatatttttc	cctaagcttt	gagcaaaagt	ttaaaaaat	acactaaaat	aatacaaaat	3177
gttaagcagt	atattagttt	gggttatata	attcatctgc	aatttataag	atgcatggcc	3237
gatgttaaat	tgcttggcaa	ttctgtaatc	attaaagtgt	ctcagtgaat	catgtcaaat	3297
gccttaaat	aactaagttg	gtgaataaaa	gtgccgatct	ggctaactct	tacaccatac	3357
atactgatag	tttttcatat	gtttcatttc	catgtgatct	ttaaaattta	gagtggaac	3417
aattttgctt	aatatggggt	acataagctt	tattttttcc	tttgttcata	atttatattc	3477
ttgaatagg	ctgtgtcaat	caagtgtatc	aactagactg	atcatagata	tgaggaaata	3537
aggccaagt	caagaccagc	ctgggcaaca	tatcgagaa	ctgtctacaa	aaaaattaaa	3597
aaaaattagc	caggcatggt	ggcgtacact	gagtagtttg	tcccagctac	tcgggaggg	3657
gaggtgggag	gatcgcttca	gcccaggagg	ttgagattgc	agtgaagccat	ggacatacca	3717
ctgcactaca	gcctaggtaa	cagcacgaga	ccccaactct	tagaaaatga	aaaggaaata	3777
tagaaatata	aaatttgctt	attatagaca	cacagtaact	cccagatatg	taccacaaaa	3837
aatgtgaaaa	gagagagaaa	tgtctaccaa	agcagtatct	tggtgtgata	attgcaagcg	3897
catagtaaaa	taattttaac	cttaatttgt	tttttagtagt	gtttagattg	aagattgagt	3957
gaaatatttt	cttggcagat	attccgtatc	tggtggaaag	ctacaatgca	atgtcgttgt	4017
agttttgcat	ggcttgcttt	ataaacaaga	ttttttctcc	ctccttttgg	gccagttttc	4077
attacgagta	actcacactt	tttgattaaa	gaacttgaaa	ttacgttatc	acttagtata	4137
attgacatta	taagagagact	atgtaacatg	caatcattag	aatcaaaatt	agtactttgg	4197
tcaaaatatt	taacaacattc	acataactgt	caaatattca	tgtaattaac	tgaattttaa	4257
accttcaact	attatgaagt	gctcgtctgt	acaatcgcta	atttactcag	tttagagtag	4317
ctacaactct	tcgatactat	catcaatatt	tgacatcttt	tccaatttgt	gtatgaaaag	4377
taaatctatt	cctgtagcaa	ctggggagtc	atatatgagg	tcaaaagacat	ataccttggt	4437
attataatat	gtatactata	ataatagctg	gttatcctga	gcaggggaaa	agggtatttt	4497
taggaaaacc	acttcaaaata	gaaagctgaa	gtacttctaa	tatactgagg	gaagtataat	4557
atgtggaaca	aactctcaac	aaaatgttta	ttgatgttga	tgaacagat	cagtttttcc	4617
atccggatta	ttattgggtc	atgattttat	atgtgaatat	gtaagatatg	ttctgcaatt	4677
ttataaatgt	tcattgtcttt	ttttaaaaaa	gggtctattg	aaattctgtg	tctccagcag	4737
gcaagaatac	ttgactaact	ctttttgtct	ctttatggta	ttttcagaat	aaagtctgac	4797
ttgtgttttt	gagattattg	gtgcctcatt	aattcagcaa	taaaggaaaa	tatgcatctc	4857
aaaaat						4863

<210> 115

<211> 5022

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> 31..33

<223> ATG

<221> misc_feature

<222> 904..906

<223> TAG

<221> polyA_signal

<222> 4995..5000

<223> AATAAA

<400> 115

```

ctgctgtccc tgggtgtcca cacgtactcc atg cgc tac ctg ctg ccc agc gtc      54
                               Met Arg Tyr Leu Leu Pro Ser Val
                               1                               5
gtg ctc ctg ggc acg gcg ccc acc tac gtg ttg gcc tgg ggg gtc tgg      102
Val Leu Leu Gly Thr Ala Pro Thr Tyr Val Leu Ala Trp Gly Val Trp
10                               15                               20
cgg ctg ctc tcc gcc ttc ctg ccc gcc cgc ttc tac caa gcg ctg gac      150
Arg Leu Leu Ser Ala Phe Leu Pro Ala Arg Phe Tyr Gln Ala Leu Asp
25                               30                               35                               40
gac cgg ctg tac tgc gtc tac cag agc atg gtg ctc ttc ttc ttc gag      198
Asp Arg Leu Tyr Cys Val Tyr Gln Ser Met Val Leu Phe Phe Phe Glu
45                               50                               55
aat tac acc ggg gtc cag cat gga gga atc tat gta aag cgc agt gcc      246
Asn Tyr Thr Gly Val Gln His Gly Gly Ile Tyr Val Lys Arg Ser Ala
60                               65                               70
aaa ttt aac gag aaa gag atg cga aac aag ttg cag agc tac gtg gac      294
Lys Phe Asn Glu Lys Glu Met Arg Asn Lys Leu Gln Ser Tyr Val Asp
75                               80                               85
gca gga act cca atg tat ctt gtg att ttt cca gaa ggt aca agg tat      342
Ala Gly Thr Pro Met Tyr Leu Val Ile Phe Pro Glu Gly Thr Arg Tyr
90                               95                               100
aat cca gag caa aca aaa gtc ctt tca gct agt cag gca ttt gct gcc      390
Asn Pro Glu Gln Thr Lys Val Leu Ser Ala Ser Gln Ala Phe Ala Ala
105                               110                               115                               120
caa cgt ggc ctt gca gta tta aaa cat gtg cta aca cca cga ata aag      438
Gln Arg Gly Leu Ala Val Leu Lys His Val Leu Thr Pro Arg Ile Lys
125                               130                               135
gca act cac gtt gct ttt gat tgc atg aag aat tat tta gat gca att      486
Ala Thr His Val Ala Phe Asp Cys Met Lys Asn Tyr Leu Asp Ala Ile
140                               145                               150
tat gat gtt acg gtg gtt tat gaa ggg aaa gac gat gga ggg cag cga      534
Tyr Asp Val Thr Val Val Tyr Glu Gly Lys Asp Asp Gly Gly Gln Arg
155                               160                               165
aga gag tca ccg acc atg acg gaa ttt ctc tgc aaa gaa tgt cca aaa      582
Arg Glu Ser Pro Thr Met Thr Glu Phe Leu Cys Lys Glu Cys Pro Lys
170                               175                               180
att cat att cac att gat cgt atc gac aaa aaa gat gtc cca gaa gaa      630
Ile His Ile His Ile Asp Arg Ile Asp Lys Lys Asp Val Pro Glu Glu
185                               190                               195                               200
caa gaa cat atg aga aga tgg ctg cat gaa cgt ttc gaa atc aaa gat      678
Gln Glu His Met Arg Arg Trp Leu His Glu Arg Phe Glu Ile Lys Asp
205                               210                               215
aag atg ctt ata gaa ttt tat gag tca cca gat cca gaa aga aga aaa      726
Lys Met Leu Ile Glu Phe Tyr Glu Ser Pro Asp Pro Glu Arg Arg Lys
220                               225                               230
aga ttt cct ggg aaa agt gtt aat tcc aaa tta agt atc aag aag act      774
Arg Phe Pro Gly Lys Ser Val Asn Ser Lys Leu Ser Ile Lys Lys Thr
235                               240                               245
tta cca tca atg ttg atc tta agt ggt ttg act gca ggc atg ctt atg      822
Leu Pro Ser Met Leu Ile Leu Ser Gly Leu Thr Ala Gly Met Leu Met
250                               255                               260
acc gat gct gga agg aag ctg tat gtg aac acc tgg ata tat gga acc      870
Thr Asp Ala Gly Arg Lys Leu Tyr Val Asn Thr Trp Ile Tyr Gly Thr
265                               270                               275                               280
cta ctt ggc tgc ctg tgg gtt act att aaa gca tag acaagtagct      916
Leu Leu Gly Cys Leu Trp Val Thr Ile Lys Ala *
285                               290
gtctccagac agtgggatgt gctacattgt ctatttttgg cggctgcaca tgacatcaaa      976
ttgtttcctg aatttattaa ggagtgtaaa taaagccttg ttgattgaag attggataat      1036

```


agaatttgtg	acpaaagctg	atatgcaatg	gtcttgggca	aacataacctg	gttgtacaac	1096
tttagcatcg	gggctgctgg	aagggtaaaa	gctaaatgga	gtttctcctg	ctctgtccat	1156
ttcctatgaa	ctaattgacaa	cttgagaagg	ctgggaggat	tgtgtatattt	gcaagtcaga	1216
tggctgcatt	tttgagcatt	aatttgcagc	gtatttcact	ttttctgtta	ttttcaattt	1276
attacaactt	gapagctcca	agctcttatt	actaaagtat	ttagtatctt	gcagctagtt	1336
aatatttcat	cttttgctta	tttctacaag	tcagtgaat	aaattgtatt	taggaagtgt	1396
caggatgttc	aaaggaaaagg	gtaaaaagtg	ttcatgggga	aaaagctctg	tttagcacat	1456
gattttattg	tattgcgtta	ttagctgatt	ttactcattt	tatatttgca	aaataaattt	1516
ctaataattt	ttgaaattgc	ttaatttgca	cacctgtac	acacagaaaa	tggtataaaa	1576
tatgagaacg	aagtttaaaa	ttgtgactct	gattcattat	agcagaactt	taaatttccc	1636
agctttttga	agatttaagc	tacgctatta	gtacttcctt	ttgtctgtgc	cataagtgct	1696
tgaaaacggt	aaggttttct	gtttttgttt	gtttttttta	tatcaaaaaga	ctcgtgtgga	1756
accttggttg	gacccaagt	tcacaagatt	tttaagggtga	tgagagcctg	cagacattct	1816
gcctagattt	actagcgtgt	gccttttgcc	tgcttctctt	tgatttcaca	gaatattcat	1876
tcagaagtcg	cgtttctgta	gtgtgggtga	ttcccactgg	gctctgggtcc	ttcccttgga	1936
tcccgtcagt	ggtgctgctc	agcggccttg	acgtagactt	gctaggaaga	aatgcagagc	1996
cagcctgtgc	tgcccacttt	cagagttgaa	ctctttaagc	ccttgtgagt	gggcttcacc	2056
agctactgca	gaggcatttt	gcatttgtct	gtgtcaagaa	gttcaccttc	tcaagccagt	2116
gaaatacaga	cttaattcgt	catgactgaa	cgaatttgtt	tatttcccat	taggtttagt	2176
ggagctacac	attaatatgt	atcgcttag	agcaagagct	gtgttccagg	aaccagatca	2236
cgatattttag	ccatggaaca	atatatccca	tgggagaaga	cctttcagtg	tgaactgttc	2296
tatttttgtg	ttataattta	aacttcgatt	tcctcatagt	cctttaagtt	gacatttctg	2356
cttactgcta	ctggattttt	gctgcagaaa	tatatcagtg	gcccacatta	aacataccag	2416
ttggatcatg	atpagcaaaa	tgaaagaaat	aatgattaaag	ggaaaaattaa	gtgactgtgt	2476
tacactgctt	cttccatgcc	agagaataaa	ctctttcaag	catcatcttt	gaagagtcgt	2536
gtgggttgaa	ttgggttgtg	tacattagaa	tgtaatgcaca	catccatgga	cactcaggat	2596
atagttggcc	taataatcgg	ggcatgggta	aaacttatga	aaatttcctc	atgctgaatt	2656
gtaattttct	cttacctgta	aagtaaaatt	tagatcaatt	ccatgtcttt	gttaagtaca	2716
gggatttaat	atattttgaa	tataatgggt	atgttctaaa	tttgaacttt	gagaggaact	2776
actgttgga	ttatgtggat	tctaactcat	tttaacaagg	tagcctgacc	tgcataagat	2836
cacttgaatg	ttagggttca	tagaactata	ctaacttctt	cacaaaagggt	ctataaaaata	2896
cagtcgttga	aaaaaatttt	gtatcaaaat	gtttggaaaa	ttagaagctt	ctccttaacc	2956
tgtattgata	ctgactgaa	tatttttcta	aaatttaagag	cgtataacct	acctgtaagt	3016
cttttccat	atcattttaa	cttttgtttg	tattattact	gatttacagc	ttagttatta	3076
atttttcttt	ataagaatgc	cgctgatgtg	catgctttta	tgtttttcag	aaaaggggtg	3136
gtttggatga	aagtaaaaaa	aaaaataaaa	tctttcactg	tctctaattg	ctgtgctgtt	3196
taacattttt	tgaccctaaa	attcaccaac	agctctccag	tacataaaaat	aggcttaagt	3256
actggccctg	cattcttcac	aatatttttc	cctaagcttt	gagcaaaagt	ttaaaaaaat	3316
acactaaaa	aatcaaaact	gttaagcagt	atattagttt	ggttatataa	attcatctgc	3376
aattttata	atgcatggcc	gatgttaatt	tgcttggcaa	ttctgtaatc	attaagtgat	3436
ctcagtga	catgtcaaat	gccttaaat	aactaagttg	gtgaataaaa	gtgccgatct	3496
ggctaactct	tacaccatac	atactgatag	tttttcatat	gtttcatttc	catgtgattt	3556
ttaaaattta	gagtggcaac	aattttgctt	aatatgggtt	acataagctt	tattttttcc	3616
tttgttcata	atttatattt	ttgaataggt	ctgtgtcaat	caagtgtatc	aactagactg	3676
atcatagata	gaaggaaata	aggccaagtt	caagaccagc	ctgggcaaca	tatcgagaac	3736
ctgtctacaa	aaaaattaaa	aaaaattagc	caggcatggt	ggcgtacact	gagtagtttg	3796
tcccagctac	tcgggagggt	gaggtgggag	gatcgcttca	gcccaggagg	ttgagattgc	3856
agttagccat	ggacatacca	ctgcactaca	gcctaggtaa	cagcacgaga	ccccactct	3916
tagaaaatga	aaaggaaata	tagaaatata	aaatttgctt	attatagaca	cacagtaact	3976
cccagatatg	taccacaaaa	aatgtgaaaa	gagagagaaa	tgtctaccaa	agcagtattt	4036
gtgtgtgata	atpgcaagcg	catagtaaaa	taattttaac	cttaatttgt	tttttagtagt	4096
gttttagattg	aaagttagtg	gaaatatttt	cttggcagat	attccgtatc	tggtggaaag	4156
ctacaatgca	atgtcgttgt	agttttgcatt	ggcttgcttt	ataaacaaga	tttttctctc	4216
ctccttttgg	gcpagttttc	attacagata	actcacactt	tttgattaaa	gaacttgaaa	4276
ttacgttatc	acttagtata	attgacatta	tatagagact	atgtaacatg	caatcattag	4336
aatcaaaatt	agtacttttg	tcaaaatatt	tacaacattc	acatacttgt	caaattattca	4396
tgtaattaac	tgaattttaaa	accttcaact	attatgaagt	gctcgtctgt	acaatcgcta	4456
atttactcag	tttagagtag	ctacaactct	tcgatactat	catcaatatt	tgacatcttt	4516
tccaatttgt	gtatgaaaag	taaatctatt	cctgtagcaa	ctggggagtc	atatatgagg	4576
tcaaagacat	ataccttgtt	attataatat	gtatactata	ataatagctg	gttatcctga	4636
gcaggggaaa	aggttatttt	taggaaaacc	acttcaaaata	gaaagctgaa	gtacttctaa	4696


```

ccccaaactct tagaaaaatga aaaggaaata tagaaatata aaatttgctt attatagaca 3876
cacagtaact cccagatatg taccacaaaa aatgtgaaaa gagagagaaa tgtctaccaa 3936
agcagtatct tgtgtgtata attgcaagcg catagtaaaa taattttaac cttaatttgt 3996
tttttagtagt gtttagattg aagattgagt gaaatatctt cttggcagat attccgtatc 4056
tggtggaaaag ctacaatgca atgtcgttgt agttttgcat ggcttgcttt ataaacaaga 4116
ttttttctcc ctctcttttg gccagttttc attacgagta actcacactt tttgattaaa 4176
gaacttgaaa ttacgttatc acttagtata attgacatta tatagagact atgtaacatg 4236
caatcattag aatcaaaaatt agtacttttg tcaaaaatatt tacaacattc acatacttgt 4296
caaatattca tgaatttaac tgaatttaaa accttcaact attatgaagt gctcgtctgt 4356
acaatcgcta atttactcag tttagagtag ctacaactct tcgatactat catcaatatt 4416
tgacatcttt tccaatttgt gtatgaaaag taaatctatt cctgtagcaa ctggggagtc 4476
atatatgagg tcaaaagacat ataccttgtt attataatat gtatactata ataatactgt 4536
gttatcctga gcaggggaaa aggttatctt taggaaaacc acttcaaata gaaagctgaa 4596
gtacttctaa tatactgagg gaagtataat atgtggaaca aactctcaac aaaatgttta 4656
ttgatgttga tgaaacagat cagtttttcc atccggatta ttattggctc atgattttat 4716
atgtgaatat gtaagatatg ttctgcaatt ttataaatgt tcatgtcttt ttttaaaaaa 4776
ggtgctattg aaattctgtg tctccagcag gcaagaatac ttgactaact ctttttgtct 4836
ctttatggta ttttcagaat aaagtctgac ttgtgttttt gagattattg gtgcctcatt 4896
aattcagcaa taaaggaaaa tatgcatctc aaaaat 4932
<210> 117
<211> 4682
<212> DNA
<213> Homo sapiens
<220>
<221> misc_feature
<222> 31..33
<223> ATG
<221> misc_feature
<222> 301..303
<223> TGA
<221> polyA_signal
<222> 4655..4660
<223> AATAAA
<400> 117
ctgctgtccc tgggtgtcca cactactcc atg cgc tac ctg ctg ccc agc gtc 54
Met Arg Tyr Leu Leu Pro Ser Val
1 5
gtg ctc ctg gcc acg gcg ccc acc tac gtg ttg gcc tgg ggg gtc tgg 102
Val Leu Leu Gly Thr Ala Pro Thr Tyr Val Leu Ala Trp Gly Val Trp
10 15 20
cgg ctg ctc tcc gcc ttc ctg ccc gcc cgc ttc tac caa gcg ctg gac 150
Arg Leu Leu Ser Ala Phe Leu Pro Ala Arg Phe Tyr Gln Ala Leu Asp
25 30 35 40
gac cgg ctg tac tgc gtc tac cag agc atg gtg ctc ttc ttc ttc gag 198
Asp Arg Leu Tyr Cys Val Tyr Gln Ser Met Val Leu Phe Phe Phe Glu
45 50 55
aat tac acc ggg gtc cag aat ttc tct gca aag aat gtc caa aaa ttc 246
Asn Tyr Thr Gly Val Gln Asn Phe Ser Ala Lys Asn Val Gln Lys Phe
60 65 70
ata ttc aca ttg atc gta tgc aca aaa aag atg tcc cag aag aac aag 294
Ile Phe Thr Leu Ile Val Ser Thr Lys Lys Met Ser Gln Lys Asn Lys
75 80 85
aac ata tga gaagatggct gcatgaacgt ttcgaaatca aagataagat 343
Asn Ile *
90
gcttatagaa ttttatgagt caccagatcc agaaagaaga aaaagatttc ctgggaaaag 403
tgtaatttcc aaattaagta tcaagaagac tttaccatca atgttgatct taagtgtttt 463
gactgcaggc atgttatga ccgatgctgg aagggaagctg tatgtgaaca cctggatata 523
tggaacccta ctgtgctgcc tgtgggttac tattaaagca tagacaagta gctgtctcca 583
gacagtggga tgtgctacat tgtctatttt tggcggctgc acatgacatc aaattgtttc 643
ctgaatttat taaggagtgt aaataaagcc ttgttgattg aagattggat aatagaattt 703

```

gtgacgaaag	ctgatatgca	atggtcttgg	gcaaacatac	ctggttgtac	aacttttagca	763
tcggggctgc	tggaagggtg	aaagctaaat	ggagtttctc	ctgctctgtc	catttcctat	823
gaactaatga	cahcttgaga	aggctgggag	gattgtgtat	tttgcaagtc	agatggctgc	883
atTTTTgagc	atTaatttgc	agcgtatttc	actttttctg	ttattttcaa	tttattacaa	943
cttgacagct	ccaagctctt	attactaaag	tatttagtat	cttgcagcta	gttaatat	1003
catcttttgc	ttatttctac	aagtcagtga	aataaattgt	atTTtaggaag	tgtcaggatg	1063
ttcaaaggaa	agggtaaaaa	gtgttcatgg	ggaaaaagct	ctgttttagca	catgatttta	1123
ttgtattgcg	ttatttagctg	attttactca	ttttatat	gcaaaataaa	tttctaata	1183
ttattgaaat	tgtttaattt	gcacaccctg	tacacacaga	aaatgggtata	aaatatgaga	1243
acgaagttta	aaattgtgac	tctgattcat	tatagcagaa	ctttaaattt	cccagctttt	1303
tgaagattta	agctacgcta	ttagtacttc	cctttgtctg	tgccataagt	gcttgaaaac	1363
gttaagggtt	tctgtttt	tttgtttt	taatatcaaa	agagtccggg	tgaaccttgg	1423
ttggacccca	agttcacaa	atttttaagg	tgatgagagc	ctgcagacat	tctgcctaga	1483
cttagtagcg	tgtgcctttt	gcctgttctc	ctttgatttc	acagaatatt	cattcagaag	1543
tcgctgttct	gtagtgtgg	ggattccccc	tggtgtctgg	tccttccctt	ggatcccgct	1603
agtggtgctg	ctcagcggct	tgacagtaga	cttgctagga	agaaatgcag	agccagcctg	1663
tgctgcccac	tttcagagtt	gaactcttta	agcccttgtg	agtgggcttc	accagctact	1723
gcagaggcat	tttgcatttg	tctgtgtcaa	gaagttccac	ttctcaagcc	agtgaataac	1783
agacttaatt	cgtcatgact	gaacgaattt	gtttatttcc	cattaggttt	agtggagcta	1843
cacattaata	tgtatcgctt	tagagcaaga	gctgtgttcc	aggaaccaga	tcacgatttt	1903
tagccatgga	acaatatatc	ccatgggaga	agacctttca	gtgtgaactg	ttctattttt	1963
gtgtataat	ttaaactctg	atttctctat	agtcctttaa	gttgacattt	ctgcttactg	2023
ctactggatt	tttgcctgag	aaatatatca	gtggccccaca	ttaaacatac	cagttggatc	2083
atgataagca	aaatgaaaga	aataatgatt	aagggaaaat	taagtgactg	tgttacactg	2143
cttctcccat	gccagagaat	aaactcttcc	aagcatcatc	tttgaagagt	cgtgtgggtg	2203
gaattgggtt	gtgtacatta	gaatgtatgc	acacatccat	ggacactcag	gatatagttg	2263
gcctaataat	cggggcatgg	gtaaaactta	tgaaaatttc	ctcatgctga	attgtaattt	2323
tctcttacct	gtaaagtaaa	atttagatca	attccatgtc	tttgttaagt	acagggattt	2383
aatataattt	gaatataatg	ggatgttctt	aaatttgaac	tttgagaggc	aatactgttg	2443
gaattatgtg	gattctaaat	cattttaaca	aggtagcctg	acctgcataa	gatcacttga	2503
atgttaggtt	tcatagaact	atactaatct	tctcacaaaa	gggtctataa	atacagtcgt	2563
tgaaaaaaat	tttgtatcaa	aatgttttga	aaattagaag	cttctcctta	acctgtattg	2623
atactgactt	gaattatttt	ctaaaattaa	gagccgtata	cctacctgta	agtcttttca	2683
catatcattt	aaacttttgt	ttgtattatt	actgatttac	agcttagtta	ttaatTTTTc	2743
tttataagaa	tgccgtcgat	gtgcctgctt	ttatgttttt	cagaaaagg	tgtgttttga	2803
tgaagtaaaa	aaaaaaaaaa	aaatctttca	ctgtctctaa	tggtgtgtgt	gtttaacatt	2863
tttgaccctt	aaaattcacc	aacagtctcc	cagtacataa	aataggctta	atgactggcc	2923
ctgcattctt	cacaatattt	ttccctaagc	tttgagcaaa	gttttaaaaa	aatacactaa	2983
ataatcaaaa	actgttaagc	agtatattag	tttgggtata	taaattcatc	tgcaatttat	3043
aagatgcatg	gccgatgtta	atttgcttgg	caattctgtg	atcattaagt	gatctcagtg	3103
aaacatgtca	aatgccttaa	atttaactaa	ttggtgaata	aaagtgccga	tctggctaac	3163
tcttacacca	tacatactga	tagtttttca	tatgtttcat	ttccatgtga	tttttaaaat	3223
ttagagtggc	aacaattttg	cttaatatgg	gttacataag	ctttattttt	tcctttgttc	3283
ataattatat	tctttgaata	ggctgtgtgc	aatcaagtga	tctaactaga	ctgatcatag	3343
atagaaggaa	ataaggccaa	gttcaagacc	agcctgggca	acatatcgag	aacctgtcta	3403
caaaaaaatt	aaaaaaaaat	agccaggcat	gggtggcgta	actgagtagt	ttgtcccagc	3463
tactcgggag	ggtgaggtgg	gaggatcgct	tcagcccagg	aggttgagat	tgcatgagc	3523
catggacata	ccactgcact	acagcctagg	taacagcacg	agaccccaac	tcttagaaaa	3583
tgaaaaggaa	atatagaaat	ataaaaattg	cttattatag	acacacagta	actcccagat	3643
atgtaccaca	aaaaatgtga	aaagagagag	aaatgtctac	caaagcagta	ttttgtgtgt	3703
ataattgcaa	gcgcatagta	aaataatttt	aaccttaatt	tgttttttagt	agtgtttaga	3763
ttgaagattg	agtgaatat	tttcttggca	gatattccgt	atctgggtga	aagctacaat	3823
gcaatgtcgt	tgtagttttg	catggcttgc	tttataaaca	agattttttc	tccttctttt	3883
tggtccagtt	ttcattacga	gtaactcaca	ctttttgatt	aaagaacttg	aaattacggt	3943
atcacttagt	ataattgaca	ttatatagag	actatgtaac	atgcaatcat	tagaatcaaa	4003
attagtactt	tggtcaaaa	atttacaaca	ttcacatact	tgtcaaatat	tcatgtaatt	4063
aactgaattt	aaaaccttca	actattatga	agtgctcgtc	tgtacaatcg	ctaatttact	4123
cagtttagag	tagctacaac	tcttcgatac	tatcatcaat	atttgacatc	ttttccaatt	4183
tgtgtatgaa	aagtaaatct	attcctgtag	caactgggga	gtcatatatg	aggtaaaaga	4243
catatacctt	gttattataa	tatgtatact	ataataatag	ctgggttatcc	tgagcagggg	4303
aaaagggttat	ttttaggaaa	accacttcaa	atagaaagct	gaagtacttc	taatatactg	4363

agggaagtat aatatgtgga acaaaactctc aacaaaatgt ttattgatgt tgatgaaaca 4423
 gatcagtttt tccatccgga ttattattgg ttcatgattt tatatgtgaa tatgtaagat 4483
 atgttctgca attttataaa tgttcattgtc tttttttaaa aaagggtgcta ttgaaattct 4543
 gtgtctccag caggcaagaa tacttgacta actctttttg tctctttatg gtattttcag 4603
 aataaagtct gacttgtgtt ttgagatta ttggtgcctc attaatcag caataaagga 4663
 aaatatgcat ctcaaaaat 4682
 <210> 118
 <211> 4558
 <212> DNA
 <213> Homo sapiens
 <220>
 <221> misc_feature
 <222> 31..33
 <223> ATG
 <221> misc_feature
 <222> 235..237
 <223> TGA
 <221> polyA_signal
 <222> 4531..4536
 <223> AATAAA
 <400> 118
 ctgctgtccc tgggtgtcca cacgtactcc atg cgc tac ctg ctg ccc agc gtc 54
 Met Arg Tyr Leu Leu Pro Ser Val
 1 5
 gtg ctc ctg ggc acg gcg ccc acc tac gtg ttg gcc tgg ggg gtc tgg 102
 Val Leu Leu Gly Thr Ala Pro Thr Tyr Val Leu Ala Trp Gly Val Trp
 10 15 20
 cgg ctg ctc tcc gcc ttc ctg ccc gcc cgc ttc tac caa gcg ctg gac 150
 Arg Leu Leu Ser Ala Phe Leu Pro Ala Arg Phe Tyr Gln Ala Leu Asp
 25 30 35 40
 gac cgg ctg tac tgc gtc tac cag agc atg gtg ctc ttc ttc gag 198
 Asp Arg Leu Tyr Cys Val Tyr Gln Ser Met Val Leu Phe Phe Glu
 45 50 55
 aat tac acc ggg gtc cag gat gct tat aga att tta tga gtcaccagat 247
 Asn Tyr Thr Gly Val Gln Asp Ala Tyr Arg Ile Leu *
 60 65
 ccagaaagaa gaaaaagatt tcttgggaaa agtggttaatt ccaaattaag tatcaagaag 307
 actttaccat caatgttgat cttaagtggg ttgactgcag gcatgcttat gaccgatgct 367
 ggaagggaagc tgtatgtgaa cacctggata tatggaaccc tacttggctg cctgtggggt 427
 actattaaag catagacaag tagctgtctc cagacagtggt gatgtgctac attgtctatt 487
 tttggcggct gcacatgaca tcaaattggt tcttgaattt attaaggagt gtaataaag 547
 ccttgttgat tgaagattgg ataatagaat ttgtgacgaa agctgatatg caatggctct 607
 gggcaaacat acctgggtgt acaactttag catcggggct gctggaaggg taaaagctaa 667
 atggagtttc tctgtctctg tccatttccct atgaactaat gacaacttga gaaggctggg 727
 aggattgtgt attttgcaag tcagatggct gcatttttga gcattaattt gcagcgattt 787
 tcacttttct tgttatttct aattttattac aacttgacag ctccaagctc ttattactaa 847
 agtatttagt atcttgcagc tagttaatat ttcattcttt gcttatttct acaagtcagt 907
 gaaataaatt gtatttagga agtgtcagga tgttcaaagg aaagggtaaa aagtgttcat 967
 ggggaaaaag ctptgttttag cacatgatgtt tattgtattg cgttattagc tgattttact 1027
 cattttatat ttgcaaaaata aatttctaatt atttattgaa attgcttaatt ttgcacaccc 1087
 tgtacacaca gaaaatggta taaaatatga gaacgaagtt taaaattgtg actctgattc 1147
 attatagcag aactttaaat ttcccagctt tttgaagatt taagctacgc tatttagtact 1207
 tccctttgtc tgtgccataa gtgcttgaaa acgttaaggt tttctgtttt gttttgtttt 1267
 tttaatatca aaagagtcgg tgtgaacctt ggttggaccc caagttcaca agatttttaa 1327
 ggtgatgaga gcctgcagac attctgccta gatttactag cgtgtgcctt ttgctgtctt 1387
 ctctttgatt tcacagaata ttcattcaga agtcgcgttt ctgtagtgtg gtggattccc 1447
 actgggctct ggtccttccc ttggatcccg tcagtgggtc tgcacagcgg cttgcacgta 1507
 gacttgctag gaagaaatgc agagccagcc tgtgctgccc actttcagag ttgaactctt 1567
 taagcccttg tgagtgggct tcaccagcta ctgcagaggc attttgcatt tgcctgtgtc 1627
 aagaagttca ccttctcaag ccagtgaat acagacttaa ttcgtcatga ctgaacgaat 1687
 ttgtttattt cccattaggt ttagtggagc tacacattaa tatgtatcgc cttagagcaa 1747

```

gagctgtgtt ccaggaacca gatcacgatt tttagccatg gaacaatata tcccatggga 1807
gaagaccttt cagtgtgaac tgttctattt ttgtgttata atttaaaactt cgatttcctc 1867
atagtccttt aagttgacat ttctgtttac tgctactgga tttttgtgc agaaatatat 1927
cagtggccca cattaaacat accagttgga tcatgataag caaaatgaaa gaaataatga 1987
ttaagggaaa attaagtgaac tgtgttacac tgcttctccc atgccagaga ataaactctt 2047
tcaagcatca tctttgaaga gtcgtgtggt gtgaattggt ttgtgtacat tagaatgtat 2107
gcacacatcc atggacactc aggatatagt tggcctaata atcggggcat gggtaaaact 2167
tatgaaaaat tccctcatgct gaattgtaat ttctctttac ctgtaaaagta aaatttagat 2227
caattccatg tctttgttaa gtacagggat ttaatatatt ttgaatataa tgggtatgtt 2287
ctaaatttga actttgagag gcaatactgt tggaattatg tggattctaa ctcattttaa 2347
caaggtagcc tgacctgcat aagatcactt gaatgttagg ttccatagaa ctataactaat 2407
cttctcaciaa aaggtctata aaatacagtc gttgaaaaaa attttgtatc aaaatgtttg 2467
gaaaattaga agcttctcct taacctgtat tgatactgac ttgaattatt ttctaaaaat 2527
aagagccgta tapctacctg taagtctttt cacatatcat ttaaaactttt gtttgtatta 2587
ttactgattt acagcttagt tattaatttt tctttataag aatgccgtcg atgtgcatgc 2647
ttttatgttt ttcagaaaag ggtgtgtttg gatgaaagta aaaaaaaaaa taaaatcttt 2707
cactgtctct aatggctgtg ctgtttaaca ttttttgacc ctaaaattca ccaacagtct 2767
cccagtagat aaaaaggctc taatgactgg cctgcattc ttcacaatat ttttccctaa 2827
gcttttagca aagtttttaa aaaatacact aaaataatca aaactgttaa gcagtatatt 2887
agtttgggta tataaattca tctgcaattt ataagatgca tggccgatgt taatttgctt 2947
ggcaattctg taatcattaa gtgactctcag tgaaacatgt caaatgcctt aaataacta 3007
agttgggtgaa taaaagtgcc gatctggcta actcttacac catacatact gatagttttt 3067
catatgtttt atttccatgt gatttttaaa atttagagtg gcaacaattt tgcctaatat 3127
gggttacata agctttattt ttctctttgt tcataattat attctttgaa taggtctgtg 3187
tcaatcaagt gatctaacta gactgatcat agatagaagg aaataaggcc aagttcaaga 3247
ccagcctggg caacatatcg agaactgtc tacaaaaaaa ttaaaaaaaa ttaggccaggc 3307
atggtggcgt acactgagta gtttgccca gctactcggg aggggtgagg gggaggatcg 3367
cttcagccca ggaggttgag attgcagtga gccatggaca taccactgca ctacagccta 3427
ggtaacagca cgagacccca actcttagaa aatgaaaagg aaatatagaa atataaaatt 3487
tgcttattat agacacacag taactcccag atatgtacca caaaaaatgt gaaaagagag 3547
agaaatgtct accaaagcag tattttgtgt gtataattgc aagcgcatag taaaataatt 3607
ttaacccttaa tttgttttta gtagtgttta gattgaagat tgagtgaatt attttcttg 3667
cagatattcc gtatctgggt gaaagctaca atgcaatgtc gttgtagttt tgcattggctt 3727
gctttataaa caagattttt tctccctcct tttgggccag ttttcattac gagtaactca 3787
cactttttga ttaaagaact tgaaattacg ttatcactta gtataattga cattatatag 3847
agactatgta acatgcaatc attagaatca aaattagtag tttggtcaaa atatttataa 3907
cattcacata ctgtgcaaat attcatgtaa ttaactgaat ttaaaacctt caactattat 3967
gaagtgtctg tctgtacaat cgctaattta ctcagtttag agtagctaca actcttcgat 4027
actatcatca atatttgaca tcttttccaa tttgtgtatg aaaagttaa ctattcctgt 4087
agcaactggg gagtcataata tgagggtcaaa gacatatacc ttgttattat aatatgtata 4147
ctataataat agctgggttat cctgagcagg ggaaaagggt atttttagga aaaccacttc 4207
aaatagaaa gctpaagtact tctaataatc tgagggaagt ataatatgtg gaacaaactc 4267
tcaacaaaat gtttattgat gttgatgaaa cagatcagtt ttccatccg gattattatt 4327
ggttcatgat ttatgatgtg aatatgtaag atatgttctg caattttata aatgttcatg 4387
tcttttttta aaaaaggtgc tattgaaatt ctgtgtctcc agcaggcaag aatacttgac 4447
taactctttt tgtctcttta tggatttttc agaataaagt ctgacttgtg ttttgagat 4507
tattggtgcc tcattaattc agcaataaag gaaaaatagc atctcaaaaa t 4558

```

<210> 119

<211> 5270

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> 31..33

<223> ATG

<221> misc_feature

<222> 229..231

<223> TAG

<221> polyA_signal

<222> 5243..5248

<223> AATAAA

<400> 119

```

ctgctgtccc tgggtgtcca cacgtactcc atg cgc tac ctg ctg ccc agc gtc      54
                               Met Arg Tyr Leu Leu Pro Ser Val
                               1       5
gtg ctc ctg ggc acg gcg ccc acc tac gtg ttg gcc tgg ggg gtc tgg      102
Val Leu Leu Gly Thr Ala Pro Thr Tyr Val Leu Ala Trp Gly Val Trp
                               10      15      20
cgg ctg ctc tcc gcc ttc ctg ccc gcc cgc ttc tac caa gcg ctg gac      150
Arg Leu Leu Ser Ala Phe Leu Pro Ala Arg Phe Tyr Gln Ala Leu Asp
25                               30      35      40
gac cgg ctg tac tgc gtc tac cag agc atg gtg ctc ttc ttc ttc gag      198
Asp Arg Leu Tyr Cys Val Tyr Gln Ser Met Val Leu Phe Phe Glu
                               45      50      55
aat tac acc ggg gtc cag aga ttg gat tgc tag attaaacttg agaaacaaac      251
Asn Tyr Thr Gly Val Gln Arg Leu Asp Ser *
                               60      65
cataaaagtg gaaggccctc tttacaata ttgctatatg gagatttggc aaaaaataaa      311
gaaaatataa tatatttagc aaatcatcaa agcacagtgt actggattgt tgctgacatc      371
ttggccatca ggcagaatgc gctaggacat gtgcgctacg tgctgaaaga aggggttaaaa      431
tggctgccat tgtatgggtg ttactttgct cagcatggag gaatctatgt aaagcgagc      491
gccaaattta acgagaaaaga gatgcgaaac aagtgtcaga gctacgtgga cgcaggaact      551
ccaatgtatc ttgtgatttt tccagaaggt acaagggtata atccagagca aacaaaagtc      611
ctttcagcta gtcaggcatt tgctgcccaa cgtggccttg cagtattaaa acatgtgcta      671
acaccacgaa taaaggcaac tcacgttgct tttgattgca tgaagaatta tttagatgca      731
atztatgatg ttacggtggt ttatgaaggg aaagacgatg gagggcagcg aagagagtca      791
ccgaccatga cgggaatttct ctgcaaagaa tgtccaaaaa ttcattatca cattgatcgt      851
atcgacaaaa aagatgtccc agaagaacaa gaacatatga gaagatggct gcatgaacgt      911
ttcgaatatc aagataagat gcttatagaa ttttatgagt caccagatcc agaaagaaga      971
aaaagatttc ctgggaaaag tgttaatttc aaattaagta tcaagaagac ttaccatca      1031
atgttgatct taagtggttt gactgcaggg atgcttatga ccgatgctgg aagggaagctg      1091
tatgtgaaca cctggatata tggaaacctc cttggctgcc tgtgggttac tattaagca      1151
tagacaagta gctgtctcca gacagtggga tgtgctacat tgtctatttt tggcggtctg      1211
acatgacatc aatttggttc ctgaatttat taaggagtgt aaataaagcc ttgttgattg      1271
aagattggat aatagaattt gtgacgaaag ctgatatgca atgggtcttg gcaaacatac      1331
ctggttgtac aacttttagca tcggggctgc tgggaagggt aaagctaaat ggagtttctc      1391
ctgctctgtc cacttcctat gaactaatga caacttgaga aggtctggag gattgtgtat      1451
tttgcaagtc agatggctgc atttttgagc attaatgtgc agcgtatttc actttttctg      1511
ttattttcaa ttattacaa cttgacagct ccaagctctt attactaaag tatttagtat      1571
cttgacgcta gttaatattt catcttttgc ttatttctac aagtcagtga aataaattgt      1631
atttaggaag tgtcaggatg ttcaaaaggaa agggtaaaaa gtgttcatgg ggaaaaagct      1691
ctgttttagc catgatttta ttgtattgct ttattagctg attttactca ttttatattt      1751
gcaaaataaa tttctaatat ttattgaaat tgcttaattt gcacaccctg tacacacaga      1811
aaatggtata aaatatgaga acgaagttaa aaattgtgac tctgattcat tatagcagaa      1871
ctttaaattt cccagctttt tgaagattta agctacgcta ttagtacttc cctttgtctg      1931
tgccataagt gcttgaaaac gtttaaggtt tctgttttgt tttgtttttt taatatcaaa      1991
agagtcggtg tgaaccttgg ttggacccca agttcacaaag atttttaagg tgatgagagc      2051
ctgcagacat tctgcctaga tttactagcg tgtgcctttt gctgcttct ctttgatttc      2111
acagaatatt cattcagaag tcgcgtttct gtagtgggtt ggattcccac tttgctctg      2171
tccttccctt ggatcccgtc agtgggtgctg ctcagcggtc tgcacgtaga cttgctagga      2231
agaaatgcag agccagcctg tgctgccac tttcagagtt gaactcttta agcccttggt      2291
agtgggcttc accagctact gcagaggcat tttgcatttg tctgtgtcaa gaagttcac      2351
ttctcaagcc agtgaatac agacttaatt cgtcatgact gaacgaattt gtttatttcc      2411
cattaggttt agtggagcta cacattaata tgtatcgct tagagcaaga gctgtgttcc      2471
aggaaccaga tcacgatttt tagccatgga acaatatatc ccatgggaga agacctttca      2531
gtgtgaactg ttctattttt gtgttataat ttaaacttcg atttcctcat agtcctttaa      2591
gttgacattt ctgcttactg ctactggatt tttgctgcag aaatatatca gtggccaca      2651
ttaaaccatac cagttggatc atgataagca aaatgaaaga aataatgatt aagggaataa      2711
taagtgactg tgttactctg cttctcccat gccagagaat aaactcttc aagcatcatc      2771
tttgaagagt cgtgtggtgt gaattggttt gtgtacatta gaatgtatgc acacatccat      2831
ggacactcag gatagattg gcctaataat cggggcatgg gtaaaactta tgaaaatttc      2891
ctcatgctga attgtaattt tctcttacct gtaaaagtaaa atttagatca attccatgct      2951

```



```

tttgtaagt acagggattt aatatatttt gaatataatg ggtatgttct aaatttgaac 3011
tttgagaggg aatactgttg gaattatgtg gatttctaact cattttaaca aggtagcctg 3071
acctgcataa gatcacttga atgttaggtt tcatagaact atactaatct tctcacaaaa 3131
gggtctataa atacagtcgt tgaaaaaaat ttgtatcaa aatgtttgga aaattagaag 3191
cttctcctta acctgtattg atactgactt gaattatttt ttaaaattaa gagccgtata 3251
cctacctgta agtcttttca catatcattt aaacttttgt ttgtattatt actgatttac 3311
agcttagtta ttaatttttc tttaataaaa tgcggtcgat gtgcattgct ttatgttttt 3371
cagaaaaggg tgtgtttgga tgaaagtaaa aaaaaaaata aaatctttca ctgtctctaa 3431
tggtgtgtgt gtttaacatt ttttgacctt aaaaattcacc aacagtcctcc cagtacataa 3491
aataggtcta atgactggcc ctgcattctt cacaattatt ttccctaagc tttgagcaaa 3551
gttttaaaaa aatacactaa aataatcaaa actgttaagc agtatattag tttgggtata 3611
taaattcatt tgcaatttat aagatgcatg gccgatgtta atttgcttgg caattctgta 3671
atcatttaagt gatctcagtg aaacatgtca aatgccttaa attaaactaag ttgttgtaata 3731
aaagtgcgga tctggctaac tcttacacca tacatactga tagtttttca tatgtttcat 3791
ttccatgtga tttttaaaat tttagagtgcc aacaattttg cttaatatgg gttacataag 3851
ctttattttt tcttttgttc ataattatat tctttgaata ggtctgtgtc aatcaagtga 3911
tctaactaga ctgatcatag atagaaggaa ataaggccaa gttcaagacc agcctgggca 3971
acatatcgag aacctgtcta caaaaaaatt aaaaaaaatt agccaggcat ggtggcgtac 4031
actgagtagt ttgtcccagc tactcgggag ggtgaggtgg gaggatcgct tcagcccagg 4091
agggttgagt tgcagtgagc catggacata ccactgcact acagcctagg taacagcacg 4151
agacccaac tcttagaaaa tgaaaaggaa atatagaaat ataaaatttg cttattatag 4211
acacacagta actcccagat atgtaccaca aaaaatgtga aaagagagag aaatgtctac 4271
caaagcagta ttttgtgtgt ataattgcaa gcgcatagta aaataatttt aaccttaatt 4331
tgtttttagt agtgtttaga ttgaagattg agtgaaatat ttctctggca gatattccgt 4391
atctgttgga aagctacaat gcaatgtcgt tgtagttttg catggcctgc tttataaaca 4451
agattttttt tccctccttt tgggcccagtt ttcatctaga gtaactcaca ctttttgatt 4511
aaagaacttg aaattacgtt atcacttagt ataattgaca ttatatagag actatgtaac 4571
atgcaatcat tagaatcaaa attagtactt tggcctcaaat atttacaaca ttcacatact 4631
tgtcaaatat tcatgtaatt aactgaattt aaaaccttca actattatga agtgctcgtc 4691
tgtacaatcg ctaatttact cagtttagag tagctacaac tcttcgatac tatcatcaat 4751
atttgacatc ttttccaatt tgtgtatgaa aagtaaatct attcctgtag caactgggga 4811
gtcatatatg aggtcaaaga catatacctt gttattataa tatgtatact ataataatag 4871
ctggttatcc tgagcagggg aaaaggttat ttttaggaaa accacttcaa atagaaagct 4931
gaagtacttc taatatactg aggggaagtat aatatgtgga acaaaactctc aacaaaatgt 4991
ttattgatgt tgatgaaaca gatcagtttt tccatccgga ttattattgg ttcattgatt 5051
tatatgtgaa tatgtaagat atgttctgca attttataaa tgttcatgtc tttttttaa 5111
aaaggtgcta ttgaaattct gtgtctccag caggcaagaa tacttgacta actctttttg 5171
tctctttatg gtattttcag aataaagtct gacttgtgtt tttgagatta ttggtgcctc 5231
attaattcag caataaaggaa aatatgcat ctcaaaaaat 5270

```

```

<210> 120
<211> 5002
<212> DNA
<213> Homo sapiens
<220>
<221> misc_feature
<222> 31..33
<223> ATG
<221> misc_feature
<222> 322..324
<223> TAA
<221> polyA_signal
<222> 4975..4980
<223> AATAAAA
<400> 120

```

```

ctgtgtgtccc tgggtgtccca cactgtactcc atg cgc tac ctg ctg ccc agc gtc 54
Met Arg Tyr Leu Leu Pro Ser Val
1 5
gtg ctc ctg ggc acg gcg ccc acc tac gtg ttg gcc tgg ggg gtc tgg 102
Val Leu Leu Gly Thr Ala Pro Thr Tyr Val Leu Ala Trp Gly Val Trp
10 15 20
cgg ctg ctc tcc gcc ttc ctg ccc gcc cgc ttc tac caa gcg ctg gac 150

```

Arg	Leu	Leu	Ser	Ala	Phe	Leu	Pro	Ala	Arg	Phe	Tyr	Gln	Ala	Leu	Asp	
25					30					35					40	
gac	cgg	ctg	tac	tgc	gtc	tac	cag	agc	atg	gtg	ctc	ttc	ttc	ttc	gag	198
Asp	Arg	Leu	Tyr	Cys	Val	Tyr	Gln	Ser	Met	Val	Leu	Phe	Phe	Phe	Glu	
			45						50						55	
aat	tac	acc	ggg	gtc	cag	ata	ttg	cta	tat	gga	gat	ttg	cca	aaa	aat	246
Asn	Tyr	Thr	Gly	Val	Gln	Ile	Leu	Leu	Tyr	Gly	Asp	Leu	Pro	Lys	Asn	
			60					65					70			
aaa	gaa	aat	ata	ata	tat	tta	gca	aat	cat	caa	agc	aca	gat	gta	tct	294
Lys	Glu	Asn	Ile	Ile	Tyr	Leu	Ala	Asn	His	Gln	Ser	Thr	Asp	Val	Ser	
		75					80					85				
tgt	gat	ttt	tcc	aga	agg	tac	aag	gta	taa	tccagagcaa	acaaaaagtcc					344
Cys	Asp	Phe	Ser	Arg	Arg	Tyr	Lys	Val	*							
		90					95									
tttcagctag	tcaggcattt															404
caccacgaat	aaaggycaact															464
tttatgatgt	tacggttggtt															524
cgaccatgat	ggaattttctc															584
tcgacaaaaa	agatgtccca															644
tcgaaatcaa	agataagatg															704
aaagatttcc	tgggaaaaagt															764
tgttgatctt	aagtggtttg															824
atgtgaacac	ctggatatat															884
agacaagtag	ctgtctccag															944
catgacatca	aatgtgtttcc															1004
agattggata	atagaatttg															1064
tggttgatca	actttagcat															1124
tgctctgtcc	atttccctatg															1184
ttgcaagtca	gatggctgca															1244
tattttcaat	ttattacaac															1304
ttgcagctag	ttaatatttc															1364
tttaggaagt	gtcaggatgt															1424
tgtttgacac	atgatttttat															1484
caaaaataaat	ttctaattatt															1544
aatggtataa	aatatgagaa															1604
tttaaatttc	ccagcttttt															1664
gccataagt	cttgaaaaag															1724
gagtcgggtg	gaaccttggt															1784
tcgacagatt	ctgcctagat															1844
cagaatattc	attcagaagt															1904
ccttcccttg	gatcccgctca															1964
gaaatcgaga	gcgagcctgt															2024
gtgggcttca	ccagctactg															2084
tctcaagcca	gtgaaataca															2144
attaggttta	ctggagctac															2204
ggaaaccagat	caggtattttt															2264
tgtgaactgt	tctatttttg															2324
ttgacatttc	tgcttactgc															2384
taaacataacc	agtgtgatca															2444
aagtgaactgt	gttacactgc															2504
ttgaagagtc	gtgtgggtgtg															2564
gacactcagg	atatagttgg															2624
tcatgctgaa	tgttaatttt															2684
ttgttaagta	cagggattta															2744
ttgagaggga	atactgttgg															2804
cctgcataag	atacattgaa															2864
gtctataaaa	tacagtctgt															2924
ttctccttaa	cctgtattga															2984
ctacctgtaa	gtcttttccac															3044
gcttagttat	taattttttct															3104
aqaaaaagqat	qtqtttqat															3164

```

ggctgtgctg tttaacattt ttgacccta aaattcacca acagtctccc agtacataaa 3224
ataggcttaa tgaactggccc tgcattcttc acaatatttt tccctaagct ttgagcaaaag 3284
ttttaaaaaa atacactaaa ataatacaaaa ctgttaagca gtatattagt ttggttatat 3344
aaattcatct gcaatttata agatgcatgg ccgatgttaa ttgcttggc aattctgtaa 3404
tcattaagtg atctcagtg aacatgtcaa atgccttaaa ttaactaagt tggngaataa 3464
aagtgccgat ctggctaact cttacaccat acatactgat agtttttcat atgtttcatt 3524
tccatgtgat ttttaaaatt tagagtggca acaattttgc ttaatatggg ttacataagc 3584
tttatttttt cctttgttca taatttatatt ctttgaatag gtctgtgtca atcaagtgat 3644
ctaactagac tgatcataga tagaaggaaa taaggccaag ttcaagacca gcctgggcaa 3704
catatcgaga acctgtctac aaaaaaatta aaaaaaatta gccaggcatg gtggcgtaaca 3764
ctgagtagtt tgtcccagct actcgggagg gtgaggtggg aggatcgctt cagcccagga 3824
ggttgagatt gpcagtggcc atggacatac cactgcacta cagcctaggt aacagcacga 3884
gaccccaact cttagaaaat gaaaaggaaa tatagaaata taaaatttgc ttattataga 3944
cacacagtaa ctcccagata tgtaccacaa aaaatgtgaa aagagagaga aatgtctacc 4004
aaagcagtat ttgtgtgta taattgcaag cgcatagtaa aataatttta accttaattt 4064
gttttttagta gtgttttagat tgaagattga gtgaaatatt ttcttggcag atattccgta 4124
tctgggtggaa agctacaatg caatgtcggt gtagttttgc atggcttgct ttataaacia 4184
gattttttct cctcctttt gggccagttt tcattacgag taactcacac tttttgatta 4244
aagaacttga aattacgtta tcacttagta taattgacat tatatagaga ctatgtaaca 4304
tgcaatcatt agaatacaaaa ttagtacttt ggtcaaaata tttaacaact tcacatactt 4364
gtcaaatatt catgtaatta actgaattta aaaccttcaa ctattatgaa gtgctcgctt 4424
gtacaatcgc taatttactc agtttagagt agctacaact cttcgatact atcatcaata 4484
tttgacatct tttccaattt gtgtatgaaa agtaaatcta ttcctgtagc aactggggag 4544
tcatatatga ggtcaaagac atataccttg ttattataat atgtatacta taataatagc 4604
tggttatcct gagcagggga aaagggttatt tttaggaaaa ccacttcaaa tagaaaagctg 4664
aagtacttct aatatactga gggaagtata atatgtggaa caaactctca acaaaatggt 4724
tattgatgtt gatgaaacag atcagttttt ccatccggat tattattggg tcatgatttt 4784
atatgtgaat atgtaagata tgttctgcaa ttttataaat gtccatgtct ttttttaaaa 4844
aagggtctat tgaattctg tgtctccagc aggcagaagt acttgactaa ctctttttgt 4904
ctctttatgg tattttcaga ataaagtctg acttgtgttt ttgagattat tgggtgcctca 4964
ttaattcagc aataaaggaa aatatgcatc tcaaaaaa 5002

```

<210> 121

<211> 4958

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> 31..33

<223> ATG

<221> misc_feature

<222> 577..579

<223> TGA

<221> polyA_signal

<222> 4931..4936

<223> AATAAA

<400> 121

```

ctgctgtccc tgggtgtcca cacgtactcc atg cgc tac ctg ctg ccc agc gtc 54
Met Arg Tyr Leu Leu Pro Ser Val
1 5
gtg ctc ctg ggc acg gcg ccc acc tac gtg ttg gcc tgg ggg gtc tgg 102
Val Leu Leu Gly Thr Ala Pro Thr Tyr Val Leu Ala Trp Gly Val Trp
10 20
cgg ctg ctc tcc gcc ttc ctg ccc gcc cgc ttc tac caa gcg ctg gac 150
Arg Leu Leu Ser Ala Phe Leu Pro Ala Arg Phe Tyr Gln Ala Leu Asp
25 30 35 40
gac cgg ctg tac tgc gtc tac cag agc atg gtg ctc ttc ttc ttc gag 198
Asp Arg Leu Tyr Cys Val Tyr Gln Ser Met Val Leu Phe Phe Phe Glu
45 50 55
aat tac acc ggg gtc cag ata ttg cta tat gga gat ttg cca aaa aat 246
Asn Tyr Thr Gly Val Gln Ile Leu Leu Tyr Gly Asp Leu Pro Lys Asn
60 65 70

```

aaa gaa aat ata ata tat tta gca aat cat caa agc aca gtt gac tgg	294
Lys Glu Asn Ile Ile Tyr Leu Ala Asn His Gln Ser Thr Val Asp Trp	
75 80 85	
att gtt gct gac atc ttg gcc atc agg cag aat gcg cta gga cat gtg	342
Ile Val Ala Asp Ile Leu Ala Ile Arg Gln Asn Ala Leu Gly His Val	
90 95 100	
cgc tac gtg ctg aaa gaa ggg tta aaa tgg ctg cca ttg tat ggg tgt	390
Arg Tyr Val Leu Lys Glu Gly Leu Lys Trp Leu Pro Leu Tyr Gly Cys	
105 110 115 120	
tac ttt gct cag cat gga gga atc tat gta aag cgc agt gcc aaa ttt	438
Tyr Phe Ala Gln His Gly Gly Ile Tyr Val Lys Arg Ser Ala Lys Phe	
125 130 135	
aac gag aaa gag atg cga aac aag ttg cag agc tac gtg gac gca gga	486
Asn Glu Lys Glu Met Arg Asn Lys Leu Gln Ser Tyr Val Asp Ala Gly	
140 145 150	
act cca aat ttc tct gca aag aat gtc caa aaa ttc ata ttc aca ttg	534
Thr Pro Asn Phe Ser Ala Lys Asn Val Gln Lys Phe Ile Phe Thr Leu	
155 160 165	
atc gta tcg aca aaa aag atg tcc cag aag aac aag aac ata tga	579
Ile Val Ser Thr Lys Lys Met Ser Gln Lys Asn Lys Asn Ile *	
170 175 180	
gaagatggct gcatgaacgt ttcgaaatca aagataagat gcttatagaa ttttatgagt	639
caccagatcc agaaagaaga aaaagatttc ctgggaaaaag tgttaattcc aaattaagta	699
tcaagaagac tttaccatca atgttgatct taagtggttt gactgcaggc atgcttatga	759
ccgatgctgg aaggaagctg tatgtgaaca cctggatata tggaaacctta cttggctgcc	819
tgtgggttac tattaaagca tagacaagta gctgtctcca gacagtggga tgtgctacat	879
tgtctatattt tggcggctgc acatgacatc aaattgtttc ctgaatttat taaggagtgt	939
aaataaagcc ttgttgattg aagattggat aatagaattt gtgacgaaag ctgatatgca	999
atggtcttgg gcaaacatcac ctggtttgtac aacttttagca tcggggctgc tgggaagggtta	1059
aaagctaaat ggagttttctc ctgctctgtc catttcctat gaactaatga caacttgaga	1119
aggctgggag gatttgttat ttgtcaagtc agatggctgc atttttgagc attaatattgc	1179
agcgtatttc actttttctg ttattttcaa tttattacaa cttgacagct ccaagctctt	1239
attactaaag tatttagtat cttgcagcta gttaaatattt catcttttgc ttatttctac	1299
aagtcagtga aataaattgt atttaggaag tgtcaggatg ttcaaaggaa agggtaaaaa	1359
gtgttcatgg ggaaaaagct ctgtttagca catgatttta ttgtattgctg ttattagctg	1419
attttactca ttttatattt gcaaaataaa tttctaatat ttattgaaat tcttaattt	1479
gcacaccctg tacacacaga aaatggtata aaatatgaga acgaagttaa aaattgtgac	1539
tctgattcat tatagcagaa ctttaaatctt cccagctttt tgaagattta agctacgcta	1599
ttagtacttc cctttgtctg tgcataagt gcttgaaaac gtttaaggttt tctgttttgt	1659
tttgtttttt taatatcaaa agagtcggtg tgaaccttgg ttggacctta agttcacaa	1719
atttttaagg tgatgagagc ctgcagacat tctgcctaga ttactagcg tgtgcctttt	1779
gcctgcttct ctttgatttc acagaatatt cattcagaag tcgcgtttct gtagtgtgg	1839
ggattccac tgggctctgg tccttccctt ggatcccgtc agtgggtgtg ctcagcggct	1899
tgcacgtaga cttgctagga agaaatgcag agccagcctg tgctgcccac tttcagagtt	1959
gaactcttta agcccttgtg agtgggcttc accagctact gcagaggcat tttgcatttg	2019
tctgtgtcaa gaagttcacc ttctcaagcc agtgaataac agacttaatt cgtcatgact	2079
gaacgaattt gtttatttcc cattaggttt agtggagcta cacattaata tgtatcgct	2139
tagagcaaga gctgtgttcc aggaaccaga tcacgatttt tagccatgga acaatatatc	2199
ccatgggaga agacctttca gtgtgaactg ttctattttt gtgttataat ttaaacttcg	2259
atttcctcat agtcctttta gttgacattt ctgcttactg ctactggatt tttgctgcag	2319
aaatatatca gtggcccaca ttaaacatcac cagttggatc atgataagca aaatgaaaga	2379
aataatgatt aagggaaaat taagtgaactg tgttacctg cttctcccat gccagagaat	2439
aaactctttc aagcatcatc tttgaagagt cgtgtggtgt gaattggttt gtgtacatta	2499
gaatgtatgc acacatccat ggacactcag gatatagttg gcctaataat cggggcatgg	2559
gtaaaactta tgaaaatttc ctcatgctga attgtaattt tctcttacct gtaaaagtaa	2619
atttagatca attccatgtc tttgttaagt acagggattt aatatatttt gaatataatg	2679
ggtatgttct aaatttgaac tttgagaggc aatactgttg gaattatgtg gattcctaact	2739
cattttaaca aggtagcctg acctgcataa gatcacttga atgttaggtt tcatagaact	2799
ataactaatct tctcacaaaa ggtctataaa atacagtcgt tgaaaaaaat tttgtatcaa	2859
aatgttttga aaattagaag cttctcctta acctgtattg atactgactt gaattatttt	2919
ctaaaattaa gagccgtata cctacctgta agtcttttca catatcattt aaacttttgt	2979

```

ttgtattatt acfgatttac agcttagtta ttaatttttc tttataagaa tgcggtcgat 3039
gtgcacgtctt ttatgttttt cagaaaaagg tgtgttttga tgaaagtaaa aaaaaaaata 3099
aaatcttttca ctgtctcttaa tggctgtgct gtttaacatt ttttgaccct aaaattcacc 3159
aacagtctcc cagtacataa aataggctta atgactggcc ctgcattctt cacaatattt 3219
ttccctaagc tttagagcaa gttttaaaaa aatacactaa aataatcaa actgttaagc 3279
agtatattag ttgggttata taaattcatc tgcaatttat aagatgcatg gccgatgtta 3339
atttgccttg caattctgta atcattaagt gatctcagtg aaacatgtca aatgccttaa 3399
attaactaag ttggtgaata aaagtgcga tctggctaac tcttacacca tacatactga 3459
tagtttttca tatgtttcat ttccatgtga tttttaaaat ttagagtggc aacaattttg 3519
cttaatatgg gttacataag ctttattttt tcctttgttc ataattatat tctttgataa 3579
ggctctgtgc aatcaagtga tctaactaga ctgatcatag atagaaggaa ataaggccaa 3639
gttcaagacc agcctgggca acatatcgag aacctgtcta caaaaaaatt aaaaaaatt 3699
agccaggcat ggtggcgtag actgagtagt ttgtcccagc tactcgggag ggtgaggtgg 3759
gaggatcgct tcagcccagg aggttgagat tgcagtggag catggacata ccactgcact 3819
acagcctaag taacagcacg agacccaac tcttagaaaa tgaaaaggaa atatagaat 3879
ataaaatttg cttattatag acacacagta actcccagat atgtaccaca aaaaatgtga 3939
aaagagagag aaatgtctac caaagcagta ttttgtgtgt ataattgcaa gcgcatagt 3999
aaataatttt aaccttaatt tgtttttagt agtgtttaga ttgaagattg agtgaaatat 4059
tttcttgcca gatattccgt atctggttga aagctacaat gcaatgtcgt tgtagttttg 4119
catggcttgc tttataaaca agattttttc tccctccttt tgggccagtt ttcattacga 4179
gtaactcaca ctttttgatt aaagaacttg aaattacgtt atcacttagt ataattgaca 4239
ttatatagag actatgtaac atgcaatcat tagaatcaaa attagtactt tggtaaaaat 4299
atttacaaca ttcacatact tgtcaaatat tcatgttaatt aactgaattt aaaaccttca 4359
actattatga agtgctcgtc tgtacaatcg ctaatttact cagttagtag tagctacaac 4419
tcttcgatac tatcatcaat atttgacatc ttttccaatt tgtgtatgaa aagtaaatct 4479
attcctgtag caactgggga gtcataatg aggtcaaaga catatacctt gttattataa 4539
tatgtatact ataataatag ctggttatcc tgagcagggg aaaaaggttat ttttaggaaa 4599
accacttcaa atagaaagct gaagtacttc taatatactg aggggaagtat aatatgttga 4659
acaaactctc aacaaaatgt ttattgatgt tgatgaaaca gatcagtttt tccatccgga 4719
ttattattgg ttcatgattt tatatgtgaa tatgtaagat atgttctgca attttataaa 4779
tgttcatgtc tttttttaaa aaaggtgcta ttgaaattct gtgtctccag caggcaagaa 4839
tacttgacta actctttttg tctctttatg gtatttttcag aataaagtct gacttgtgtt 4899
tttgagatta ttggtgcctc attaattcag caataaagga aaatatgcat ctcaaaaat 4958
<210> 122
<211> 5094
<212> DNA
<213> Homo sapiens
<220>
<221> misc_feature
<222> 31..33
<223> ATG
<221> misc_feature
<222> 976..978
<223> TAG
<221> polyA_signal
<222> 5067..5072
<223> AATAAA
<400> 122
ctgctgtccc tgggtgctcca cacgtactcc atg cgc tac ctg ctg ccc agc gtc 54
Met Arg Tyr Leu Leu Pro Ser Val
1 5
gtg ctc ctg ggc acg gcg ccc acc tac gtg ttg gcc tgg ggg gtc tgg 102
Val Leu Leu Gly Thr Ala Pro Thr Tyr Val Leu Ala Trp Gly Val Trp
10 15 20
cgg ctg ctc tcc gcc ttc ctg ccc gcc cgc ttc tac caa gcg ctg gac 150
Arg Leu Leu Ser Ala Phe Leu Pro Ala Arg Phe Tyr Gln Ala Leu Asp
25 30 35 40
gac cgg ctg tac tgc gtc tac cag agc atg gtg ctc ttc ttc ttc gag 198
Asp Arg Leu Tyr Cys Val Tyr Gln Ser Met Val Leu Phe Phe Glu
45 50 55
aat tac acc ggg gtc cag ata ttg cta tat gga gat ttg cca aaa aat 246

```

Asn Tyr Thr Gly Val Gln Ile Leu Leu Tyr Gly Asp Leu Pro Lys Asn
 60 65 70
 aaa gaa aat ata ata tat tta gca aat cat caa agc aca gtt gac tgg 294
 Lys Glu Asn Ile Ile Tyr Leu Ala Asn His Gln Ser Thr Val Asp Trp
 75 80 85
 att gtt gct gac atc ttg gcc atc agg cag aat gcg cta gga cat gtg 342
 Ile Val Ala Asp Ile Leu Ala Ile Arg Gln Asn Ala Leu Gly His Val
 90 95 100
 cgc tac gtg ctg aaa gaa ggg tta aaa tgg ctg cca ttg tat ggg tgt 390
 Arg Tyr Val Leu Lys Glu Gly Leu Lys Trp Leu Pro Leu Tyr Gly Cys
 105 110 115 120
 tac ttt gct cag cat gga gga atc tat gta aag cgc agt gcc aaa ttt 438
 Tyr Phe Ala Gln His Gly Gly Ile Tyr Val Lys Arg Ser Ala Lys Phe
 125 130 135
 aac gag aaa gag atg cga aac aag ttg cag agc tac gtg gac gca gga 486
 Asn Glu Lys Glu Met Arg Asn Lys Leu Gln Ser Tyr Val Asp Ala Gly
 140 145 150
 act cca atg tat ctt gtg att ttt cca gaa ggt aca agg tat aat cca 534
 Thr Pro Met Tyr Leu Val Ile Phe Pro Glu Gly Thr Arg Tyr Asn Pro
 155 160 165
 gag caa aca aaa gtc ctt tca gct agt cag gca ttt gct gcc caa cgt 582
 Glu Gln Thr Lys Val Leu Ser Ala Ser Gln Ala Phe Ala Ala Gln Arg
 170 175 180
 ggg aaa gac gat gga ggg cag cga aga gag tca ccg acc atg acg gaa 630
 Gly Lys Asp Asp Gly Gln Arg Arg Glu Ser Pro Thr Met Thr Glu
 185 190 195 200
 ttt ctc tgc aaa gaa tgt cca aaa att cat att cac att gat cgt atc 678
 Phe Leu Cys Lys Glu Cys Pro Lys Ile His Ile His Ile Asp Arg Ile
 205 210 215
 gac aaa aaa gat gtc cca gaa gaa caa gaa cat atg aga aga tgg ctg 726
 Asp Lys Lys Asp Val Pro Glu Glu Gln Glu His Met Arg Arg Trp Leu
 220 225 230
 cat gaa cgt ttc gaa atc aaa gat aag atg ctt ata gaa ttt tat gag 774
 His Glu Arg Phe Glu Ile Lys Asp Lys Met Leu Ile Glu Phe Tyr Glu
 235 240 245
 tca cca gat cca gaa aga aga aaa aga ttt cct ggg aaa agt gtt aat 822
 Ser Pro Asp Pro Glu Arg Arg Lys Arg Phe Pro Gly Lys Ser Val Asn
 250 255 260
 tcc aaa tta agt atc aag aag act tta cca tca atg ttg atc tta agt 870
 Ser Lys Leu Ser Ile Lys Lys Thr Leu Pro Ser Met Leu Ile Leu Ser
 265 270 275 280
 ggt ttg act gca ggc atg ctt atg acc gat gct gga agg aag ctg tat 918
 Gly Leu Thr Ala Gly Met Leu Met Thr Asp Ala Gly Arg Lys Leu Tyr
 285 290 295
 gtg aac acc tgg ata tat gga acc cta ctt ggc tgc ctg tgg gtt act 966
 Val Asn Thr Trp Ile Tyr Gly Thr Leu Gly Cys Leu Trp Val Thr
 300 305 310
 att aaa gca tag acaagtagct gtctccagac agtgggatgt gctacattgt 1018
 Ile Lys Ala
 315
 ctatttttgg cggctgcaca tgacatcaaa ttgtttcctg aattttattaa ggagtgtaaa 1078
 taaagccttg ttgattgaag attggataat agaatttctg acgaaagctg atatgcaatg 1138
 gtcttgggca aacatacctg gttgtacaac tttagcatcg gggctgctgg aagggtaaaa 1198
 gctaaatgga gtttctcctg ctctgtccat ttcctatgaa ctaatgacaa cttgagaagg 1258
 ctgggaggat tgtgtatttt gcaagtcaga tggctgcatt tttgagcatt aatttgcagc 1318
 gtatttcaact ttttctgtta ttttcaattt attacaactt gacagctcca agctcttatt 1378
 actaaagtat ttagtatctt gcagctagtt aatatttcat cttttgctta tttctacaag 1438
 tcagtgaat aaattgtatt taggaagtgt caggatgttc aaaggaaagg gtaaaaaagtg 1498
 ttcatgggga aaaagctctg tttagcacat gattttattg tattgcgtta ttagctgatt 1558
 ttactcattt tatatttgca aaataaattt ctaataattta ttgaaattgc ttaatttgca 1618
 caccctgtac acacagaaaa tgggtataaaa tatgagaacg aagtttaaaa ttgtgactct 1678

gattcattat	agcagaactt	taaatttccc	agcttttttga	agattttaagc	tacgctatta	1738
gtacttccct	ttgtctgtgc	cataagtgc	tgaaaaacgtt	aaggttttct	gttttgtttt	1798
gttttttttaa	tatcaaaaaga	gtcgggtgtga	accttggttg	gaccccaagt	tcacaagatt	1858
tttaagggtga	tgagagcctg	cagacattct	gcctagattt	actagcgtgt	gccttttggc	1918
tgcttctctt	tgatttcaca	gaatattcat	tcagaagtcg	cgtttctgta	gtgtgggtgga	1978
ttcccactgg	gctctgggtcc	ttcccttgga	tcccgctcagt	ggtgctgctc	agcggccttg	2038
acgtagactt	gctaggaaga	aatgcagagc	cagcctgtgc	tgcccacttt	cagagttgaa	2098
ctctttaagc	ccttgtgagt	gggcttcacc	agctactgca	gaggcatttt	gcatttgtct	2158
gtgtcaagaa	gttcaccttc	tcaagccagt	gaaatacaga	cttaattcgt	catgactgaa	2218
cgaatttggt	tatttcccat	taggtttagt	ggagctacac	attaatatgt	atcgcccttag	2278
agcaagagct	gtgttccagg	aaccagatca	cgatttttag	ccatggaaca	atatatccca	2338
tgggagaaga	cctttcagtg	tgaactgttc	tatttttgtg	ttataattta	aacttcgatt	2398
tcctcatagt	cctttaagtt	gacatttctg	cttactgcta	ctggattttt	gctgcagaaa	2458
tatatcagtg	gpcacattta	aacataccag	ttggatcatg	ataagcaaaa	tgaagaaatt	2518
aatgatttaag	ggaaaattaa	gtgactgtgt	tacactgctt	ctcccatgcc	agagaataaa	2578
ctctttcaag	catcatcttt	gaagagtcgt	gtgggtgtgaa	ttgggtttgtg	tacattagaa	2638
tgatgcaca	catccatgga	cactcaggat	atagttggcc	taataatcgg	ggcatgggta	2698
aaactttatga	aaatttctct	atgctgaatt	gtaattttct	cttacctgta	aagtataaatt	2758
tagatcaatt	ccatgtcttt	gttaagtaca	gggattttaat	atattttgaa	tataatgggt	2818
atgttctaaa	tttgaacttt	gagaggcaat	actgttggaa	ttatgtggat	tctaactcat	2878
tttaacaagg	tagcctgacc	tgcataagat	cacttgaatg	ttaggtttca	tagaactata	2938
ctaattctct	cacaaaagg	ctataaaata	cagtcggttg	aaaaaatttt	gtatcaaaat	2998
gtttggaaaa	ttagaagctt	ctccttaacc	tgtattgata	ctgacttgaa	ttattttcta	3058
aaattaagag	ccgtatacct	acctgtaagt	cttttcacat	atcattttaa	cttttgtttg	3118
tattattact	gatttcacag	ttagttatta	atttttcttt	ataagaatgc	cgctgatgtg	3178
catgctttta	tgtttttcag	aaaagggtgt	gtttggatga	aagtaaaaaa	aaaaataaaa	3238
tctttcactg	tctctaattg	ctgtgctgtt	taacattttt	tgaccctaaa	attcaccaac	3298
agtctccag	tacataaaat	aggcttaatg	actggccctg	cattcttcac	aatatttttc	3358
cctaagcttt	gagcaaagt	ttaaaaaat	acactaaaat	aatcaaaact	gttaagcagt	3418
atattagttt	ggttatataa	attcatctgc	aatttataag	atgcatggcc	gatgttaatt	3478
tgcttgga	ttctgtaatc	attaagtgat	ctcagtga	catgtcaa	gccttaaat	3538
aactaagttg	gtgaataaaa	gtgccgatct	ggctaactct	tacaccatac	atactgatag	3598
tttttcata	gtttcatttc	catgtgattt	ttaaaattta	gagtggcaac	aattttgctt	3658
aatatgggtt	acataagctt	tattttttcc	ttgttcata	attatattct	ttgaataggt	3718
ctgtgtcaat	caagtgatct	aactagactg	atcatagata	gaaggaaata	aggccaaagt	3778
caagaccagc	ctgggcaaca	tatcgagaac	ctgtctacaa	aaaaattaaa	aaaaatttagc	3838
caggcatggt	ggcgtaact	gagtagtttg	tcccagctac	tcgggagggg	gaggtgggag	3898
gatcgcttca	gccaggagg	ttgagattgc	agtgagccat	ggacatacca	ctgactaca	3958
gcctaggtaa	cagcacgaga	ccccaaactct	tagaaaaatga	aaaggaaata	tagaaatata	4018
aaatttgctt	attatagaca	cacagtaact	cccagatatg	taccacaaaa	aatgtgaaaa	4078
gagagagaaa	tgtctaccaa	agcagtattt	tgtgtgtata	attgcaagcg	catagtaaaa	4138
taatttttaac	cttaatttgt	tttttagtagt	gttttagattg	aagatttagt	gaaatatttt	4198
cttggcagat	atccggtatc	tggtggaaag	ctacaatgca	atgtcgttgt	agttttgcat	4258
ggcttgcttt	ataaacaaga	ttttttctcc	ctccttttgg	gccagttttc	attacgagta	4318
actcacactt	tttgattaaa	gaacttgaaa	ttacgttatc	acttagtata	attgacatta	4378
tatagagact	atgtaacatg	caatcattag	aatcaaaatt	agtactttgg	tcaaaatatt	4438
tacaacattc	acatacttgt	caaataattca	tgttaattac	tgaattttaa	accttcaact	4498
attatgaagt	gctcgtctgt	acaatcgcta	atttactcag	tttagagtag	ctacaactct	4558
tcgatactat	catcaatatt	tgacatcttt	tccaatttgt	gtatgaaaag	taaactctatt	4618
cctgtagcaa	ctggggagtc	atatatgagg	tcaaagacat	ataccttggt	attataatat	4678
gtatactata	ataatagctg	gttatcctga	gcaggggaaa	aggttatttt	taggaaaacc	4738
acttcaaaata	gaaagctgaa	gtacttctaa	tatactgagg	gaagtataat	atgtggaaca	4798
aactctcaac	aaaatgttta	ttgatgttga	tgaaacagat	cagtttttcc	atccggatta	4858
ttatttggtt	atgattttat	atgtgaatat	gtaagatatg	ttctgcaatt	ttataaatgt	4918
tcatgtcttt	ttttaaaaaa	ggtgctattg	aaattctgtg	tctccagcag	gcaagaatac	4978
ttgactaact	ctttttgtct	ctttatggta	ttttcagaat	aaagtctgac	ttgtgttttt	5038
gagattattg	gtgcctcatt	aattcagcaa	taaaggaaaa	tatgcatctc	aaaaat	5094

<210> 123

<211> 5049

<212> DNA

<213> Homo sapiens

250	255	260	
agt ggt ttg act gca ggc atg ctt atg acc gat gct gga agg aag ctg			870
Ser Gly Leu Thr Ala Gly Met Leu Met Thr Asp Ala Gly Arg Lys Leu			
265	270	275	280
tat gtg aac acc tgg ata tat gga acc cta ctt ggc tgc ctg tgg gtt			918
Tyr Val Asn Thr Trp Ile Tyr Gly Thr Leu Leu Gly Cys Leu Trp Val			
285	290	295	
act att aaa gca tag acaagtagct gtctccagac agtgggatgt gctacattgt			973
Thr Ile Lys Ala *			
300			
ctatttttgg cggctgcaca tgacatcaaa ttgtttcctg aattttattaa ggagtgtaaa			1033
taaagccttg ttgattgaag attggataat agaattttgtg acgaaagctg atatgcaatg			1093
gtcttgggca aacataacctg gttgtacaac tttagcatcg gggctgctgg aagggtaaaa			1153
gctaaatgga gtttctcctg ctctgtccat ttccctatgaa ctaatgacaa cttgagaagg			1213
ctgggaggat tgtgtatttt gcaagtcaga tggctgcatt tttgagcatt aatttgcagc			1273
gtatttcaact ttttctgtta ttttcaattt attacaactt gacagctcca agctcttatt			1333
actaaagtat ttagtatctt gcagctagtt aatatttcat cttttgtcta tttctacaag			1393
tcagtgaat aaattgtatt taggaagtgt caggatgttc aaaggaaaagg gtaaaaagtg			1453
ttcatgggga aaaagctctg ttttagcacat gatttttattg tattgctgta ttagctgatt			1513
ttactcattt tatatttgca aaataaattt ctaatattta ttgaaattgc ttaatttgca			1573
cacctgtac acacagaaaa tgggtataaaa tatgagaacg aagtttaaaa ttgtgactct			1633
gattcattat agcagaactt taaatttccc agctttttga agattttaagc tacgctatta			1693
gtacttccct ttgtctgtgc cataagtgtc tgaaaacgtt aagggttttct gttttgtttt			1753
gtttttttaa tatcaaaaga gtcggtgtga accttggttg gaccccaagt tcacaagatt			1813
tttaagggtga tgagagcctg cagacattct gcctagattt actagcgtgt gccttttgcc			1873
tgcttctctt tgatttcaca gaatttccat tcagaagtgc cgtttctgta gtgtgggtga			1933
ttcccactgg gctctggtcc ttcccttggga tcccgtcagt ggtgctgctc agcggcttgc			1993
acgtagactt gctaggaaga aatgcagagc cagcctgtgc tgcccacttt cagagttgaa			2053
ctctttaagc ccttgtgagt gggcttcacc agctactgca gaggcatttt gcatttgcct			2113
gtgtcaagaa gttcaccttc tcaagccagt gaaatacaga cttaattcgt catgactgaa			2173
cgaatttgtt tatttcccat taggtttagt ggagctacac attaatatgt atcgccttag			2233
agcaagagct gtgttccagg aaccagatca cgatttttag ccatggaaca atatatccca			2293
tgggagaaga cctttcagtg tgaactgttc tatttttgtg ttataattta aacttcgatt			2353
ttctcatagt cctttaagtt gacatttctg ctactgtcta ctggattttt gtcgagaaa			2413
tatatcagtg gccacatta aacataccag ttggatcatg ataagcaaaa tgaaagaaat			2473
aatgattaa ggaataataa gtgactgtgt tacactgtct ctcccatgcc agagaataaa			2533
ctctttcaag catcatcttt gaagagtcgt gtggtgtgaa ttggtttgtg tacattagaa			2593
tgtatgcaca catccatgga cactcaggat atagtggcc taataatcgg ggcatgggtg			2653
aaacttatga aaatttctc atgctgaatt gtaattttct cttacctgta aagtaaaatt			2713
tagatcaatt ccatgtcttt gttaagtaca gggatttaat atattttgaa tataatgggt			2773
atgttctaaa tttgaacttt gagaggcaat actgttgga ttatgtggat tctaactcat			2833
tttaacaagg tagcctgacc tgcataagat cacttgaatg ttagggtttca tagaactata			2893
ctaatttctt caaaaaggt ctataaaata cagtcgttga aaaaaatttt gtatcaaaat			2953
gtttggaataa ttagaagctt ctcttaacc tgtattgata ctgacttgaa ttattttcta			3013
aaattaagag ccgtatacct acctgtaagt cttttcacat atcattttaa cttttgtttg			3073
tattattact gatttacagc ttagttatta atttttctt ataagaatgc cgtcgatgtg			3133
catgctttta tgtttttcag aaaagggtgt gtttggatga aagtaaaaaa aaaaataaaa			3193
tccttctactg tctctaattg ctgtgctgtt taacattttt tgaccctaaa attcaccac			3253
agtctcccag tacataaaat aggtttaatg actggccctg cattcttcac aatatttttc			3313
cctaagcttt gagcaaggtt ttaaaaaaat acactaaaat aatcaaaact gtaagcagt			3373
atattagttt ggttatataa attcatctgc aatttataag atgcatggcc gatgttaatt			3433
tgcttggcaa ttctgtaatc attaagtgat ctgagtgaaa catgtcaaat gccttaatt			3493
aactaagttg gtgaataaaa gtgccgatct ggctaactct tacaccatac atactgatag			3553
tttttcatat gtttcatttc catgtgattt ttaaaaattta gagtggcaac aattttgctt			3613
aatatgggtt acataagctt tattttttcc tttgttcata attatattct ttgaataggt			3673
ctgtgtcaat caagtgtatc aactagactg atcatagata gaaggaaata aggccaaagt			3733
caagaccagc ctgggcaaac tatcgagaac ctgtctacaa aaaaattaaa aaaaattagc			3793
caggcatggt gggtacact gagtagtttg tcccagctac tggggagggt gaggtgggag			3853
gatcgcttca gccaggagg ttgagattgc agtgagccat ggacatacca ctgcactaca			3913
gcctaggtaa cagcacgaga ccccaactct tagaaaatga aaaggaaata tagaaatata			3973
aaatttgctt attatagaca cacagtaact cccagatatg taccacaaaa aatgtgaaaa			4033

gagagagaaa tgtctaccaa agcagtat tttgtgtata attgcaagcg catagtaaaa 4093
 taattttaac ctttaattgt ttttagtagt gtttagattg aagattgagt gaaatatttt 4153
 cttggcagat attccgtatc tgggtgaaaag ctacaatgca atgtcgttgt agttttgcat 4213
 ggcttgcttc ataaacaaga tttttctcc ctcccttttg gccagttttc attacgagta 4273
 actcacactt tttgattaaa gaacttgaaa ttacgttatc acttagtata attgacatta 4333
 tatagagact atgtaacatg caatcattag aatcaaaaatt agtacttttg tcaaaatatt 4393
 tacaacattc acatacttgt caaatattca tgtaattaac tgaattttaa accttcaact 4453
 attatgaagt gctcgtctgt acaatcgcta atttactcag ttttagagtag ctacaactct 4513
 tgcatactat catcaatatt tgacatcttt tccaatttgt gtatgaaaag taaatctatt 4573
 cctgtagcaa ctggggagtc atatatgagg tcaaagacat ataccttgtt attataatat 4633
 gtatactata ataatactg gttatcttga gcaggggaaa aggttatttt taggaaaacc 4693
 acttcaaata gaaaagctgaa gtacttctaa tatactgagg gaagtataat atgtgggaaca 4753
 aactctcaac aaaaatgttta ttgatgttga tgaacacagat cagtttttcc atccggatta 4813
 ttattggttc atgattttat atgtgaatat gtaagatatg ttctgcaatt ttataaatgt 4873
 tcatgtcttt ttttaaaaaa ggtgctattg aaattctgtg tctccagcag gcaagaatac 4933
 ttgactaact ctttttgtct ctttatggta ttttcagaat aaagtctgac ttgtgttttt 4993
 gagattattg gtgcctcatt aattcagcaa taaaggaaaa tatgcatctc aaaaaat 5049
 <210> 124
 <211> 5324
 <212> DNA
 <213> Homo sapiens
 <220>
 <221> misc_feature
 <222> 31..33
 <223> ATG
 <221> misc_feature
 <222> 586..588
 <223> TAA
 <221> polyA_signal
 <222> 5297..5302
 <223> AATAAA
 <400> 124
 ctgctgtccc tgggtgtcca cagctactcc atg cgc tac ctg ctg ccc agc gtc 54
 Met Arg Tyr Leu Leu Pro Ser Val
 1 5
 gtg ctc ctg ggc acg gcg ccc acc tac gtg ttg gcc tgg ggg gtc tgg 102
 Val Leu Leu Gly Thr Ala Pro Thr Tyr Val Leu Ala Trp Gly Val Trp
 10 15 20
 cgg ctg ctc tcc gcc ttc ctg ccc gcc cgc ttc tac caa gcg ctg gac 150
 Arg Leu Leu Ser Ala Phe Leu Pro Ala Arg Phe Tyr Gln Ala Leu Asp
 25 30 35 40
 gac cgg ctg tac tgc gtc tac cag agc atg gtg ctc ttc ttc ttc gag 198
 Asp Arg Leu Tyr Cys Val Tyr Gln Ser Met Val Leu Phe Phe Glu
 45 50 55
 aat tac acc ggg gtc cag ata ttg cta tat gga gat ttg cca aaa aat 246
 Asn Tyr Thr Gly Val Gln Ile Leu Leu Tyr Gly Asp Leu Pro Lys Asn
 60 65 70
 aaa gaa aat ata ata tat tta gca aat cat caa agc aca gtt gac tgg 294
 Lys Glu Asn Ile Ile Tyr Leu Ala Asn His Gln Ser Thr Val Asp Trp
 75 80 85
 att gtt gct gac atc ttg gcc atc agg cag aat gcg cta gga cat gtg 342
 Ile Val Ala Asp Ile Leu Ala Ile Arg Gln Asn Ala Leu Gly His Val
 90 95 100
 cgc tac gtg ctg aaa gaa ggg tta aaa tgg ctg cca ttg tat ggg tgt 390
 Arg Tyr Val Leu Lys Glu Gly Leu Lys Trp Leu Pro Leu Tyr Gly Cys
 105 110 115 120
 tac ttt gct cag cat gga gga atc tat gta aag cgc agt gcc aaa ttt 438
 Tyr Phe Ala Gln His Gly Gly Ile Tyr Val Lys Arg Ser Ala Lys Phe
 125 130 135
 aac gag aaa gag atg cga aac aag ttg cag agc tac gtg gac gca gga 486
 Asn Glu Lys Glu Met Arg Asn Lys Leu Gln Ser Tyr Val Asp Ala Gly

act cca atg	140	145	150	
Thr Pro Met Tyr Leu Val Ile Phe Pro Glu Gly Thr Arg Tyr Asn Pro				534
gag caa aca aaa gtc ctt tca gct agt cag gca ttt gct gcc caa cgt	155	160	165	582
Glu Gln Thr Lys Val Leu Ser Ala Ser Gln Ala Phe Ala Ala Gln Arg				
ggc taa agcagtcctc ctgagtagtt aggactacag acatacacgt gccaccgcgc	170	175	180	638
Gly *				
185				
ccagctccgt gttctctttg tttccctgcc tctgtctctt ccacttatct ttgcatggca				698
ggccttgagc tattaaaaa tgtgctaaca ccacgaataa aggcaactca cgttgctttt				758
gattgcatga agaattattt agatgcaatt tatgatgtta cgggtggtta tgaaggga				818
gacgatggag ggcagcgaag agagtcaccg accatgacgg aatttctctg caaagaatgt				878
ccaaaaattc atattcacat tgatcgtatc gacaaaaaag atgtcccaga agaacaagaa				938
catatgagaa gatggctgca tgaacgtttc gaaatcaaag ataagatgct tatgaaat				998
tatgagtcac cagatccaga aagaagaaaa agatttctctg ggaaaagtgt taattccaaa				1058
ttaagtatca agaagacttt accatcaatg ttgatcttaa gtggtttgac tgcaggcatg				1118
cttatgaccg atgctggaag gaagctgtat gtgaacacct ggatataatg aacctactt				1178
ggctgcctgt gggttactat taaagcatag acaagtagct gtctccagac agtgggatgt				1238
gctacattgt ctatttttgg cggtgcaca tgacatcaaa ttgtttcctg aatttattaa				1298
ggagtgtaaa taaagccttg ttgattgaag attggataat agaatttggc acgaaagctg				1358
atatgcaatg gtcttgggca aacatacctg gttgtacaac tttagcatcg gggctgctgg				1418
aagggttaaa gctaaatgga gtttctcctg ctctgtccat ttcttatgaa ctaatgacaa				1478
cttgagaagg ctgggaggat tgtgtatttt gcaagtcaga ttgctgcatt tttagcatt				1538
aatttgcagc gtatttcaact ttttctgtta ttttcaattt attacaactt gacagctcca				1598
agctcttatt actaaagtat tttagtatctt gcagctagtt aatatttcat cttttgctta				1658
tttctacaag tcagtgaat aaattgtatt taggaagtgt caggatgttc aaaggaaagg				1718
gtaaaaagtg ttcatgggga aaaagctctg tttagcacat gattttattg tattgcttta				1778
ttagctgatt ttactcattt tatatttgca aaataaattt ctaataattt ttgaaattgc				1838
ttaatgtgca caccctgtac acacagaaaa tggatataaa tatgagaacg aagtttaaaa				1898
ttgtgactct gattcattat agcagaactt taaatttccc agctttttga agatttaagc				1958
tacgctatta gtacttccct ttgtctgtgc cataagtgtc tgaacacgtt aaggttttct				2018
gttttgtttt gtttttttaa tatcaaaaaga gtcggtgtga accttgggtg gaccccaagt				2078
tcacaagatt ttttaaggta tgagagcctg cagacattct gcctagattt actagcgtgt				2138
gccttttggc tgcttctctt tgatttcaca gaatttcat tcagaagtcg cgtttctgta				2198
gtgtggtgga ttcccactgg gctctggtcc ttcccttgga tcccgtcagt ggtgctgctc				2258
agcggcttg acgtagactt gctaggaaga aatgcagagc cagcctgtgc tgcccacttt				2318
cagagttgaa ctctttaagc ccttgtgagt gggcttcacc agctactgca gaggcatttt				2378
gcatttgtct gtgtcaagaa gttcaccttc tcaagccagt gaaatacaga ctttaattcg				2438
catgactgaa cgaatttgtt tatttcccat taggtttagt ggagctacac attaatatgt				2498
atcgccctag agcaagagct gtgttcagg aaccagatca cgatttttag ccatggaaca				2558
atatatccca tgggagaaga cctttcagtg tgaactgttc tatttttgtg ttataattta				2618
aacttcgatt tctcatagtt cctttaagtt gacattttct cttactgcta ctggattttt				2678
gtgagcaaaa tatatcagtg gccacatta aacataccag ttggatcatg ataagcaaaa				2738
tgaagaaat aatgattaag ggaataatga gtgactgtgt tacactgctt ctcccagcc				2798
agagaataaa ctctttcaag catcatcttt gaagagtcgt gtggtgtgaa ttggtttgtg				2858
tacattagaa tgtatgcaca catccatgga cactcaggat atagttggcc taataatcgg				2918
ggcatgggta aaacttatga aaatttctc atgctgaatt gtaattttct ctacctgta				2978
aagtaaaatt tagatcaatt ccatgtcttt gtttaagtaca gggatttaat atattttgaa				3038
tataatgggt atgttctaaa tttagaactt gagaggcaat actgttgga ttatgtggat				3098
tctaactcat ttttaacaagg tagcctgacc tgcataagat cacttgaatg ttaggtttca				3158
tagaactata ctaactcttct cacaataagg ctataaaaata cagtcgttga aaaaaatttt				3218
gtatcaaaat gtttggaataa ttagaagctt ctcttaacc tgtattgata ctgacttgaa				3278
ttattttcta aaatttaagag ccgtataacct acctgtaagt cttttcacat atcattttaa				3338
cttttgtttg tattattact gatttacagc tttagttatta atttttcttt ataagaaatg				3398
cgctgatgtg catgctttta tgtttttcag aaaagggtgt gtttggatga aagtaaaaaa				3458
aaaaataaaa tctttcactg tctctaattg ctgtgctgtt taacattttt tgaccctaaa				3518
attcaccaac agtctcccag tacataaaaat aggccttaatg actggccctg cattcttcac				3578
aatatttttc cctaagcttt gagcaaatgt ttaaaaaaat acactaaaat aatcaaaact				3638
gttaagcagt atattagttt ggttatataa attcatctgc aatttataag atgcatggcc				3698

```

gatgttaatt tgcttgcaa ttctgtaatc attaagtgat ctcaagtcaa catgtcaaat 3758
gccttaaatt aactaagttg gtgaataaaa gtgccgatct ggctaactct tacaccatac 3818
atactgatat tttttcatat gtttcatttc catgtgattt ttaaaattta gagtggcaac 3878
aattttgctt aatatgggtt acataagctt tattttttcc tttgttcata attatattct 3938
ttgaataggt ctgtgtcaat caagtgtatc aactagactg atcatagata gaaggaaata 3998
aggccaagtt caagaccagc ctgggcaaca tatcgagaac ctgtctacaa aaaaattaaa 4058
aaaaattagc caggcatggt ggcgtaactt gtagtatttg tcccagctac tcgggagggg 4118
gaggtgggag gatcgcttca gcccaggagg ttgagattgc agtgagccat ggacatacca 4178
ctgcactaca gcctaggtaa cagcacgaga ccccaactct tagaaaatga aaaggaaata 4238
tagaaaatata aaatttgctt attatagaca cacagtaact cccagatatg taccacaaaa 4298
aatgtgaaaa gagagagaaa tgtctaccaa agcagtattt tgtgtgtata attgcaagcg 4358
catagtaaaa taattttaac cttaatttgt ttttagtagt gtttagattg aagattgagt 4418
gaaatatttt ctgggcagat attccgtatc tgggtgaaaag ctacaatgca atgtcgttgt 4478
agttttgcat ggcttgcttt ataaacaaga ttttttctcc ctctttttgg gccagttttc 4538
attacgagta actcacactt tttgattaaa gaacttgaaa ttacgttata acttagtata 4598
attgacatta tatagagact atgtaacatg caatcattag aatcaaaatt agtactttgg 4658
tcaaaaatatt tacaacattc acatacttgt caaatattca tgaatttaac tgaatttaaa 4718
acctcaactt attatgaagt gctcgcttgt acaatcgcta atttactcag tttagagttag 4778
ctacaactct tcgatactat catcaatatt tgacatcttt tccaatttgt gtatgaaaag 4838
taaatctatt cctgtagcaa ctggggagtc atatatgagg tcaaagacat atacctgtgt 4898
attataatat gtatactata ataatactgt gttatcctga gcaggggaaa aggttatttt 4958
taggaaaacc acttcaaaata gaaagctgaa gtactttctaa tatactgagg gaagtataat 5018
atgtggaaca aactctcaac aaaatgttta ttgatgttga tgaacagat cagtttttcc 5078
atccggatta ttattgggtt atgattttat atgtgaatat gtaagatatg ttctgcaatt 5138
ttataaatgt tcatgtcttt ttttaaaaaa ggtgctattg aaattctgtg tctccagcag 5198
gcaagaatac ttgactaact ctttttgtct ctttatggta ttttcagaat aaagtctgac 5258
ttgtgttttt gagattattg gtgcctcatt aattcagcaa taaaggaaaa tatgcatctc 5318
aaaaat 5324

```

<210> 125

<211> 77

<212> PRT

<213> Homo sapiens

<400> 125

```

Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
1          5          10          15
Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
20          25          30
Ala Arg Phe Tyr Gln Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln
35          40          45
Ser Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln Leu Thr
50          55          60
Gly Leu Leu Leu Thr Ser Trp Pro Ser Gly Arg Met Arg
65          70          75

```

<210> 126

<211> 238

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> 98..103

<223> Box II

<400> 126

```

Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
1          5          10          15
Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
20          25          30
Ala Arg Phe Tyr Gln Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln
35          40          45
Ser Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln His Gly
50          55          60
Gly Ile Tyr Val Lys Arg Ser Ala Lys Phe Asn Glu Lys Glu Met Arg

```


Ser Pro Asp Pro Glu Arg Arg Lys Arg Phe Pro Gly Lys Ser Val Asn
 225 230 235 240
 Ser Lys Leu Ser Ile Lys Lys Thr Leu Pro Ser Met Leu Ile Leu Ser
 245 250 255
 Gly Leu Thr Ala Gly Met Leu Met Thr Asp Ala Gly Arg Lys Leu Tyr
 260 265 270
 Val Asn Thr Trp Ile Tyr Gly Thr Leu Leu Gly Cys Leu Trp Val Thr
 275 280 285
 Ile Lys Ala
 290
 <210> 128
 <211> 261
 <212> PRT
 <213> Homo sapiens
 <220>
 <221> SITE
 <222> 68..73
 <223> Box II
 <221> SITE
 <222> 119..127
 <223> Box III
 <400> 128
 Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
 1 5 10 15
 Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
 20 25 30
 Ala Arg Phe Tyr Gln Ala Leu Asp Arg Leu Tyr Cys Val Tyr Gln
 35 40 45
 Ser Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln Met Tyr
 50 55 60
 Leu Val Ile Phe Pro Glu Gly Thr Arg Tyr Asn Pro Glu Gln Thr Lys
 65 70 75 80
 Val Leu Ser Ala Ser Gln Ala Phe Ala Ala Gln Arg Gly Leu Ala Val
 85 90 95
 Leu Lys His Val Leu Thr Pro Arg Ile Lys Ala Thr His Val Ala Phe
 100 105 110
 Asp Cys Met Lys Asn Tyr Leu Asp Ala Ile Tyr Asp Val Thr Val Val
 115 120 125
 Tyr Glu Gly Lys Asp Asp Gly Gln Arg Arg Glu Ser Pro Thr Met
 130 135 140
 Thr Glu Phe Leu Cys Lys Glu Cys Pro Lys Ile His Ile His Ile Asp
 145 150 155 160
 Arg Ile Asp Lys Lys Asp Val Pro Glu Glu Gln Glu His Met Arg Arg
 165 170 175
 Trp Leu His Glu Arg Phe Glu Ile Lys Asp Lys Met Leu Ile Glu Phe
 180 185 190
 Tyr Glu Ser Pro Asp Pro Glu Arg Arg Lys Arg Phe Pro Gly Lys Ser
 195 200 205
 Val Asn Ser Lys Leu Ser Ile Lys Lys Thr Leu Pro Ser Met Leu Ile
 210 215 220
 Leu Ser Gly Leu Thr Ala Gly Met Leu Met Thr Asp Ala Gly Arg Lys
 225 230 235 240
 Leu Tyr Val Asn Thr Trp Ile Tyr Gly Thr Leu Leu Gly Cys Leu Trp
 245 250 255
 Val Thr Ile Lys Ala
 260
 <210> 129
 <211> 90
 <212> PRT
 <213> Homo sapiens
 <400> 129

```

Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
1           5           10           15
Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
20           25           30
Ala Arg Phe Tyr Gln Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln
35           40           45
Ser Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln Asn Phe
50           55           60
Ser Ala Lys Asn Val Gln Lys Phe Ile Phe Thr Leu Ile Val Ser Thr
65           70           75           80
Lys Lys Met Ser Gln Lys Asn Lys Asn Ile
85           90

```

<210> 130

<211> 68

<212> PRT

<213> Homo sapiens

<400> 130

```

Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
1           5           10           15
Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
20           25           30
Ala Arg Phe Tyr Gln Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln
35           40           45
Ser Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln Asp Ala
50           55           60
Tyr Arg Ile Leu
65

```

<210> 131

<211> 66

<212> PRT

<213> Homo sapiens

<400> 131

```

Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
1           5           10           15
Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
20           25           30
Ala Arg Phe Tyr Gln Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln
35           40           45
Ser Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln Arg Leu
50           55           60
Asp Ser
65

```

<210> 132

<211> 97

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> 81..83

<223> Box I

<400> 132

```

Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
1           5           10           15
Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
20           25           30
Ala Arg Phe Tyr Gln Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln
35           40           45
Ser Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln Ile Leu
50           55           60
Leu Tyr Gly Asp Leu Pro Lys Asn Lys Glu Asn Ile Ile Tyr Leu Ala
65           70           75           80

```

Asn His Gln Ser Thr Asp Val Ser Cys Asp Phe Ser Arg Arg Tyr Lys
 85 90 95

Val

<210> 133

<211> 182

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> 81..83

<223> Box I

<400> 133

Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
 1 5 10 15
 Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
 20 25 30
 Ala Arg Phe Tyr Gln Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln
 35 40 45
 Ser Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln Ile Leu
 50 55 60
 Leu Tyr Gly Asp Leu Pro Lys Asn Lys Glu Asn Ile Ile Tyr Leu Ala
 65 70 75 80
 Asn His Gln Ser Thr Val Asp Trp Ile Val Ala Asp Ile Leu Ala Ile
 85 90 95
 Arg Gln Asn Ala Leu Gly His Val Arg Tyr Val Leu Lys Glu Gly Leu
 100 105 110
 Lys Trp Leu Pro Leu Tyr Gly Cys Tyr Phe Ala Gln His Gly Gly Ile
 115 120 125
 Tyr Val Lys Arg Ser Ala Lys Phe Asn Glu Lys Glu Met Arg Asn Lys
 130 135 140
 Leu Gln Ser Tyr Val Asp Ala Gly Thr Pro Asn Phe Ser Ala Lys Asn
 145 150 155 160
 Val Gln Lys Phe Ile Phe Thr Leu Ile Val Ser Thr Lys Lys Met Ser
 165 170 175
 Gln Lys Asn Lys Asn Ile
 180

<210> 134

<211> 315

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> 81..83

<223> Box I

<221> SITE

<222> 160..165

<223> Box II

<400> 134

Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
 1 5 10 15
 Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
 20 25 30
 Ala Arg Phe Tyr Gln Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln
 35 40 45
 Ser Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln Ile Leu
 50 55 60
 Leu Tyr Gly Asp Leu Pro Lys Asn Lys Glu Asn Ile Ile Tyr Leu Ala
 65 70 75 80
 Asn His Gln Ser Thr Val Asp Trp Ile Val Ala Asp Ile Leu Ala Ile
 85 90 95
 Arg Gln Asn Ala Leu Gly His Val Arg Tyr Val Leu Lys Glu Gly Leu

100 105 110
 Lys Trp Leu Pro Leu Tyr Gly Cys Tyr Phe Ala Gln His Gly Gly Ile
 115 120 125
 Tyr Val Lys Arg Ser Ala Lys Phe Asn Glu Lys Glu Met Arg Asn Lys
 130 135 140
 Leu Gln Ser Tyr Val Asp Ala Gly Thr Pro Met Tyr Leu Val Ile Phe
 145 150 155 160
 Pro Glu Gly Thr Arg Tyr Asn Pro Glu Gln Thr Lys Val Leu Ser Ala
 165 170 175
 Ser Gln Ala Phe Ala Ala Gln Arg Gly Lys Asp Asp Gly Gly Gln Arg
 180 185 190
 Arg Glu Ser Pro Thr Met Thr Glu Phe Leu Cys Lys Glu Cys Pro Lys
 195 200 205
 Ile His Ile His Ile Asp Arg Ile Asp Lys Lys Asp Val Pro Glu Glu
 210 215 220
 Gln Glu His Met Arg Arg Trp Leu His Glu Arg Phe Glu Ile Lys Asp
 225 230 235 240
 Lys Met Leu Ile Glu Phe Tyr Glu Ser Pro Asp Pro Glu Arg Arg Lys
 245 250 255
 Arg Phe Pro Gly Lys Ser Val Asn Ser Lys Leu Ser Ile Lys Lys Thr
 260 265 270
 Leu Pro Ser Met Leu Ile Leu Ser Gly Leu Thr Ala Gly Met Leu Met
 275 280 285
 Thr Asp Ala Gly Arg Lys Leu Tyr Val Asn Thr Trp Ile Tyr Gly Thr
 290 295 300
 Leu Leu Gly Cys Leu Trp Val Thr Ile Lys Ala
 305 310 315
 <210> 135
 <211> 300
 <212> PRT
 <213> Homo sapiens
 <220>
 <221> SITE
 <222> 81..83
 <223> Box I
 <221> SITE
 <222> 160..165
 <223> Box II
 <400> 135
 Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
 1 5 10 15
 Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
 20 25 30
 Ala Arg Phe Tyr Gln Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln
 35 40 45
 Ser Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln Ile Leu
 50 55 60
 Leu Tyr Gly Asp Leu Pro Lys Asn Lys Glu Asn Ile Ile Tyr Leu Ala
 65 70 75 80
 Asn His Gln Ser Thr Val Asp Trp Ile Val Ala Asp Ile Leu Ala Ile
 85 90 95
 Arg Gln Asn Ala Leu Gly His Val Arg Tyr Val Leu Lys Glu Gly Leu
 100 105 110
 Lys Trp Leu Pro Leu Tyr Gly Cys Tyr Phe Ala Gln His Gly Gly Ile
 115 120 125
 Tyr Val Lys Arg Ser Ala Lys Phe Asn Glu Lys Glu Met Arg Asn Lys
 130 135 140
 Leu Gln Ser Tyr Val Asp Ala Gly Thr Pro Met Tyr Leu Val Ile Phe
 145 150 155 160
 Pro Glu Gly Thr Arg Tyr Asn Pro Glu Gln Thr Lys Val Leu Ser Ala
 165 170 175

Ser Gln Ala Phe Ala Ala Gln Arg Glu Phe Leu Cys Lys Glu Cys Pro
 180 185 190
 Lys Ile His Ile His Ile Asp Arg Ile Asp Lys Lys Asp Val Pro Glu
 195 200 205
 Glu Gln Glu His Met Arg Arg Trp Leu His Glu Arg Phe Glu Ile Lys
 210 215 220
 Asp Lys Met Leu Ile Glu Phe Tyr Glu Ser Pro Asp Pro Glu Arg Arg
 225 230 235 240
 Lys Arg Phe Pro Gly Lys Ser Val Asn Ser Lys Leu Ser Ile Lys Lys
 245 250 255
 Thr Leu Pro Ser Met Leu Ile Leu Ser Gly Leu Thr Ala Gly Met Leu
 260 265 270
 Met Thr Asp Ala Gly Arg Lys Leu Tyr Val Asn Thr Trp Ile Tyr Gly
 275 280 285
 Thr Leu Leu Gly Cys Leu Trp Val Thr Ile Lys Ala
 290 295 300

<210> 136

<211> 185

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> 81..83

<223> Box I

<221> SITE

<222> 160..165

<223> Box II

<400> 136

Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
 1 5 10 15
 Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
 20 25 30
 Ala Arg Phe Tyr Gln Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln
 35 40 45
 Ser Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln Ile Leu
 50 55 60
 Leu Tyr Gly Asp Leu Pro Lys Asn Lys Glu Asn Ile Ile Tyr Leu Ala
 65 70 75 80
 Asn His Gln Ser Thr Val Asp Trp Ile Val Ala Asp Ile Leu Ala Ile
 85 90 95
 Arg Gln Asn Ala Leu Gly His Val Arg Tyr Val Leu Lys Glu Gly Leu
 100 105 110
 Lys Trp Leu Pro Leu Tyr Gly Cys Tyr Phe Ala Gln His Gly Gly Ile
 115 120 125
 Tyr Val Lys Arg Ser Ala Lys Phe Asn Glu Lys Glu Met Arg Asn Lys
 130 135 140
 Leu Gln Ser Tyr Val Asp Ala Gly Thr Pro Met Tyr Leu Val Ile Phe
 145 150 155 160
 Pro Glu Gly Thr Arg Tyr Asn Pro Glu Gln Thr Lys Val Leu Ser Ala
 165 170 175
 Ser Gln Ala Phe Ala Ala Gln Arg Gly
 180 185

<210> 137

<211> 19

<212> DNA

<213> Homo Sapiens

<220>

<221> misc_binding

<222> 1..19

<223> amplification oligonucleotide PGIASel3

<400> 137

accgggggtcc agttgactg	19
<210> 138	
<211> 17	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> misc_binding	
<222> 1..17	
<223> amplification oligonucleotide PG1ASe14	
<400> 138	
cggggtccag catggag	17
<210> 139	
<211> 16	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> misc_binding	
<222> 1..16	
<223> amplification oligonucleotide PG1ASe15	
<400> 139	
ccgggggtcca ggcctt	16
<210> 140	
<211> 16	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> misc_binding	
<222> 1..16	
<223> amplification oligonucleotide PG1ASe16	
<400> 140	
cggggtccag gccttg	16
<210> 141	
<211> 21	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> misc_binding	
<222> 1..21	
<223> amplification oligonucleotide PG1ASe17	
<400> 141	
accgggggtcc agaatttctc t	21
<210> 142	
<211> 19	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> misc_binding	
<222> 1..19	
<223> amplification oligonucleotide PG1ASe18	
<400> 142	
cggggtccag gatgcttat	19
<210> 143	
<211> 23	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> misc_binding	
<222> 1..23	
<223> amplification oligonucleotide PG1ASe24	
<400> 143	
aatcatcaaa gcacagcatg gag	23

<210> 144
 <211> 28
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> misc_binding
 <222> 1..28
 <223> amplification oligonucleotide PG1ASe25
 <400> 144
 caaatcatca aagcacagat gtatcttg 28
 <210> 145
 <211> 20
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> misc_binding
 <222> 1..20
 <223> amplification oligonucleotide PG1ASe26
 <400> 145
 atcaaagcac aggccttgca 20
 <210> 146
 <211> 26
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> misc_binding
 <222> 1..26
 <223> amplification oligonucleotide PG1ASe27
 <400> 146
 agcaaatcat caaagcacag aatttc 26
 <210> 147
 <211> 28
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> misc_binding
 <222> 1..28
 <223> amplification oligonucleotide PG1ASe28
 <400> 147
 atcatcaaag cacaggatgc ttatagaa 28
 <210> 148
 <211> 31
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> misc_binding
 <222> 1..31
 <223> amplification oligonucleotide PG1ASe35
 <400> 148
 gtgttacttt gctcagatgt atcttgat t 31
 <210> 149
 <211> 23
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> misc_binding
 <222> 1..23
 <223> amplification oligonucleotide PG1ASe36
 <400> 149
 tactttgctc aggccttgca gta 23
 <210> 150

```

<211> 27
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..27
<223> amplification oligonucleotide PG1ASe37
<400> 150
gggtgttact ttgctcagaa tttctct
27
<210> 151
<211> 29
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..29
<223> amplification oligonucleotide PG1ASe38
<400> 151
ggtgttactt tgctcaggat gcttataga
29
<210> 152
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..20
<223> amplification oligonucleotide PG1ASe46
<400> 152
caggaactcc agccttgag
20
<210> 153
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..23
<223> amplification oligonucleotide PG1ASe47
<400> 153
caggaactcc aaatttctct gca
23
<210> 154
<211> 25
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..25
<223> amplification oligonucleotide PG1ASe48
<400> 154
cgcaggaact ccagatgctt ataga
25
<210> 155
<211> 22
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..22
<223> amplification oligonucleotide PG1ASe57
<400> 155
ctgccaacg tgaatttctc tg
22
<210> 156
<211> 22

```

<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..22
<223> amplification oligonucleotide PG1ASe58
<400> 156
gccaacgtg gatgcttata ga 22
<210> 157
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..23
<223> amplification oligonucleotide PG1ASe68
<400> 157
cgaccatgac gggatgctta tag 23
<210> 158
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..19
<223> amplification oligonucleotide PG1ASe1X
<400> 158
ccggggtcca gagattgga 19
<210> 159
<211> 26
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..26
<223> amplification oligonucleotide PG1ASeX2
<400> 159
aaagtggaag gccctcttta acaata 26
<210> 160
<211> 25
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..25
<223> amplification oligonucleotide PG1Aelb3
<400> 160
gccctcttta acattgactg gattg 25
<210> 161
<211> 24
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..24
<223> amplification oligonucleotide PG1Aelb4
<400> 161
gccctcttta acacatggag gaat 24
<210> 162
<211> 28
<212> DNA

```

<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..28
<223> amplification oligonucleotide PG1Ae1b5
<400> 162
ggccctcttt aacaatgtat cttgtgat
<210> 163
<211> 25
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..25
<223> amplification oligonucleotide PG1Ae1b6
<400> 163
ggccctcttta acagccttgc agtat
<210> 164
<211> 25
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..25
<223> amplification oligonucleotide PG1Ae1b7
<400> 164
ggccctcttt aacaaatttc tctgc
<210> 165
<211> 28
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..28
<223> amplification oligonucleotide PG1Ae1b8
<400> 165
gaaggccctc tttacagat gcttatag
<210> 166
<211> 26
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..26
<223> amplification oligonucleotide PG1Ae3b4
<400> 166
atgctggatt atagcatgga ggaatc
<210> 167
<211> 31
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..31
<223> amplification oligonucleotide PG1Ae3b5
<400> 167
caaaatgctg gattatagat gtatcttgtg a
<210> 168
<211> 23
<212> DNA
<213> Homo Sapiens

```

```

<220>
<221> misc_binding
<222> 1..23
<223> amplification oligonucleotide PG1Ae3b6
<400> 168
tgctggatta taggccttgc agt
<210> 169
<211> 28
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..28
<223> amplification oligonucleotide PG1Ae3b7
<400> 169
tgctggatta tagaatttct ctgcaaag
<210> 170
<211> 30
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..30
<223> amplification oligonucleotide PG1Ae3b8
<400> 170
ccaaaatgct ggattatagg atgcttatag
<210> 171
<211> 21
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..21
<223> amplification oligonucleotide PG1Ae5b6
<400> 171
tatctttgca tggcagcctt g
<210> 172
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..23
<223> amplification oligonucleotide PG1Ae5b7
<400> 172
ctttgcatgg caaatttctc tgc
<210> 173
<211> 27
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..27
<223> amplification oligonucleotide PG1Ae5b8
<400> 173
ttatctttgc atggcagatg cttatag
<210> 174
<211> 20
<212> DNA
<213> Homo Sapiens
<220>

```

<221> misc_binding
 <222> 1..20
 <223> amplification oligonucleotide PG1Ae56b
 <400> 174
 ctgcccacg tgggaaagac 20
 <210> 175
 <211> 21
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> misc_binding
 <222> 1..21
 <223> amplification oligonucleotide PG1Ae46b
 <400> 175
 gcaggaaactc caggaaagac g 21
 <210> 176
 <211> 25
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> misc_binding
 <222> 1..25
 <223> amplification oligonucleotide PG1Ae36b
 <400> 176
 tgttactttg ctcagggaagac gacga 25
 <210> 177
 <211> 22
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> misc_binding
 <222> 1..22
 <223> amplification oligonucleotide PG1Ae26b
 <400> 177
 atcaaagcac agggaaagac ga 22
 <210> 178
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> misc_binding
 <222> 1..19
 <223> amplification oligonucleotide PG1Ae16b
 <400> 178
 ccgggggtcca gggaaagac 19
 <210> 179
 <211> 56520
 <212> DNA
 <213> Homo sapiens
 <220>
 <221> exon
 <222> 2001..2216
 <223> exon1
 <221> exon
 <222> 18196..18265
 <223> exon2
 <221> exon
 <222> 23716..23831
 <223> exon3
 <221> exon
 <222> 25570..25659

<223> exon4
<221> exon
<222> 34668..34758
<223> exon5
<221> exon
<222> 40685..40843
<223> exon6
<221> exon
<222> 48067..48190
<223> exon7
<221> exon
<222> 50179..54519
<223> exon8
<221> polyA_signal
<222> 54493..54498
<223> AATAAA
<221> primer_bind
<222> 1991..2008
<223> upstream amplification primer 5-63
<221> primer_bind
<222> 2505..2525
<223> downstream amplification primer 5-63 , complement
<221> primer_bind
<222> 4091..4111
<223> downstream amplification primer 99-622
<221> primer_bind
<222> 4528..4546
<223> upstream amplification primer 99-622 , complement
<221> primer_bind
<222> 5475..5495
<223> downstream amplification primer 99-621
<221> primer_bind
<222> 5927..5947
<223> upstream amplification primer 99-621 , complement
<221> primer_bind
<222> 8127..8144
<223> downstream amplification primer 99-619
<221> primer_bind
<222> 8560..8578
<223> upstream amplification primer 99-619 , complement
<221> primer_bind
<222> 11622..11639
<223> upstream amplification primer 4-76
<221> primer_bind
<222> 12018..12037
<223> downstream amplification primer 4-76 , complement
<221> primer_bind
<222> 11930..11947
<223> upstream amplification primer 4-77
<221> primer_bind
<222> 12339..12358
<223> downstream amplification primer 4-77 , complement
<221> primer_bind
<222> 12915..12932
<223> upstream amplification primer 4-71
<221> primer_bind
<222> 13317..13334
<223> downstream amplification primer 4-71 , complement
<221> primer_bind
<222> 13216..13233
<223> upstream amplification primer 4-72

<221> primer_bind
<222> 13617..13636
<223> downstream amplification primer 4-72 , complement
<221> primer_bind
<222> 13547..13564
<223> upstream amplification primer 4-73
<221> primer_bind
<222> 13962..13981
<223> downstream amplification primer 4-73 , complement
<221> primer_bind
<222> 15994..16011
<223> downstream amplification primer 99-610
<221> primer_bind
<222> 16463..16480
<223> upstream amplification primer 99-610 , complement
<221> primer_bind
<222> 17304..17324
<223> downstream amplification primer 99-609
<221> primer_bind
<222> 17814..17832
<223> upstream amplification primer 99-609 , complement
<221> primer_bind
<222> 18008..18027
<223> upstream amplification primer 4-90
<221> primer_bind
<222> 18423..18442
<223> downstream amplification primer 4-90 , complement
<221> primer_bind
<222> 18699..18716
<223> downstream amplification primer 99-607
<221> primer_bind
<222> 19164..19182
<223> upstream amplification primer 99-607 , complement
<221> primer_bind
<222> 22589..22609
<223> downstream amplification primer 99-602
<221> primer_bind
<222> 23111..23129
<223> upstream amplification primer 99-602 , complement
<221> primer_bind
<222> 25098..25118
<223> downstream amplification primer 99-600
<221> primer_bind
<222> 25657..25674
<223> upstream amplification primer 99-600 , complement
<221> primer_bind
<222> 26537..26557
<223> downstream amplification primer 99-598
<221> primer_bind
<222> 27022..27040
<223> upstream amplification primer 99-598 , complement
<221> primer_bind
<222> 32262..32281
<223> downstream amplification primer 99-592
<221> primer_bind
<222> 32823..32841
<223> upstream amplification primer 99-592 , complement
<221> primer_bind
<222> 34215..34233
<223> upstream amplification primer 99-217
<221> primer_bind

<222> 34624..34644
<223> downstream amplification primer 99-217 , complement
<221> primer_bind
<222> 34473..34491
<223> upstream amplification primer 5-47
<221> primer_bind
<222> 34916..34936
<223> downstream amplification primer 5-47 , complement
<221> primer_bind
<222> 34702..34722
<223> downstream amplification primer 99-589
<221> primer_bind
<222> 35182..35200
<223> upstream amplification primer 99-589 , complement
<221> primer_bind
<222> 39591..39611
<223> upstream amplification primer 99-12899
<221> primer_bind
<222> 39971..39991
<223> downstream amplification primer 99-12899 , complement
<221> primer_bind
<222> 40531..40549
<223> upstream amplification primer 4-12
<221> primer_bind
<222> 40932..40950
<223> downstream amplification primer 4-12 , complement
<221> primer_bind
<222> 40629..40649
<223> downstream amplification primer 99-582
<221> primer_bind
<222> 41058..41078
<223> upstream amplification primer 99-582 , complement
<221> primer_bind
<222> 45729..45746
<223> downstream amplification primer 99-576
<221> primer_bind
<222> 46186..46203
<223> upstream amplification primer 99-576 , complement
<221> primer_bind
<222> 47879..47896
<223> upstream amplification primer 4-13
<221> primer_bind
<222> 48217..48236
<223> downstream amplification primer 4-13 , complement
<221> primer_bind
<222> 48902..48922
<223> upstream amplification primer 99-12903
<221> primer_bind
<222> 49331..49351
<223> downstream amplification primer 99-12903 , complement
<221> primer_bind
<222> 49830..49848
<223> upstream amplification primer 5-56
<221> primer_bind
<222> 50271..50290
<223> downstream amplification primer 5-56 , complement
<221> primer_bind
<222> 50172..50189
<223> upstream amplification primer 4-61
<221> primer_bind
<222> 50573..50591

<223> downstream amplification primer 4-61 , complement
<221> primer_bind
<222> 50541..50560
<223> upstream amplification primer 4-62
<221> primer_bind
<222> 50940..50959
<223> downstream amplification primer 4-62 , complement
<221> primer_bind
<222> 50555..50572
<223> upstream amplification primer 4-63
<221> primer_bind
<222> 50964..50983
<223> downstream amplification primer 4-63 , complement
<221> primer_bind
<222> 50774..50792
<223> upstream amplification primer 4-64
<221> primer_bind
<222> 51183..51202
<223> downstream amplification primer 4-64 , complement
<221> primer_bind
<222> 51146..51165
<223> upstream amplification primer 4-65
<221> primer_bind
<222> 51479..51496
<223> downstream amplification primer 4-65 , complement
<221> primer_bind
<222> 51593..51610
<223> upstream amplification primer 4-67
<221> primer_bind
<222> 29734..29744
<223> upstream amplification primer 4-67 , complement
<221> primer_bind
<222> 51167..51185
<223> upstream amplification primer 5-50
<221> primer_bind
<222> 51667..51687
<223> downstream amplification primer 5-50 , complement
<221> primer_bind
<222> 51387..51403
<223> upstream amplification primer 5-71
<221> primer_bind
<222> 51826..51843
<223> downstream amplification primer 5-71 , complement
<221> primer_bind
<222> 51772..51789
<223> upstream amplification primer 5-30
<221> primer_bind
<222> 52199..52217
<223> downstream amplification primer 5-30 , complement
<221> primer_bind
<222> 51850..51867
<223> upstream amplification primer 5-58
<221> primer_bind
<222> 52382..52400
<223> downstream amplification primer 5-58 , complement
<221> primer_bind
<222> 52507..52527
<223> upstream amplification primer 5-53
<221> primer_bind
<222> 52997..53017
<223> downstream amplification primer 5-53 , complement

```

<221> primer_bind
<222> 52703..52721
<223> upstream amplification primer 5-60
<221> primer_bind
<222> 53142..53162
<223> downstream amplification primer 5-60 , complement
<221> primer_bind
<222> 53001..53018
<223> upstream amplification primer 5-68
<221> primer_bind
<222> 53521..53538
<223> downstream amplification primer 5-68 , complement
<221> primer_bind
<222> 53459..53476
<223> upstream amplification primer 5-66
<221> primer_bind
<222> 53920..53940
<223> downstream amplification primer 5-66 , complement
<221> primer_bind
<222> 54202..54220
<223> upstream amplification primer 5-62
<221> primer_bind
<222> 54681..54701
<223> downstream amplification primer 5-62 , complement
<400> 179
gtggatctgt gactgttcgc aggaagagag gagcgggagc aggacagaca ataactgata      60
gtcaggagct gggtttgag ataaagaggg aacaagagaa agttaagttc tgtgttttca      120
tggcaaacat tgcacaaaag tttacaactt cgtgactaac agtaatctgg ggtgattcac      180
aacaaattta cacataaaca catatttact gactttatac acagcaatcc taacgtgaac      240
acagaacctg ctttatcttt tcgcacactg ttctagtgtg gagatgtctg gtctcagtta      300
aagaaagcat aaggagcatt agttgtgcac actgtccaca cccgtgactt ttttccacca      360
gtactaaacc tagtgcttct tacagtacag ggcaatgaca gccacagaaa gagagaagct      420
ccttttactg tgtaatgctt cctgctggcc ttcaaatact tgttacttga gagatctcca      480
ttcacctggc tttgtcccca aaggtcatca tctaccaatg atgttggtat ttgatgttaa      540
tcattgtataa agaaagtagc taccatctcg gccctgatta gaacttccca ctgaaatacc      600
gtcctgccta aaggtagcac aggtttccat tatggtggtg gtggggaggg ggccgggaata      660
tatatatata tatatatata tatatatatg gtaaagcatt cggcattctt ttaaagtaca      720
actatccttg aaaaggggta catattaaac cttttttacc acagccaaag gggaggagaa      780
agatccaaaa gtccgtgtga tctgctttaa catcaataaa acagtatcc acccttcgta      840
gcttttagtg aaggctacaa aagtatgctt tttatggatt acacatgtgc acgcaactac      900
tttaattact acagaaaaaa acgaggctcc ttattaaaaa aaaatcagaa acaagtccaa      960
cagactctga ggaaatgaag caagagtga tttctgaaaag gtctaataaa cagtatggaa      1020
atatccttgt gggattgttc ttacagctatg cataaacatg taattatcat cttactgtg      1080
atggggaaaa acacggaccc taattctgaa acaccctggt agcgagagac gggcaggagg      1140
ggctgctgcg cactcagagc ggaggctgag gaggcggcgt ccccttgcaa aggactggca      1200
gtgagcagat ggggacactc gagctgcccc gcgacctggg ccgagctgcc tacaacctgg      1260
gcccagggtg ctgcaagaat tagacctcgg ataacgttaa caccacttt ctcactgctc      1320
taattgtgtg catcccgcg cccaggggct tgtgagcagc aggtgcgctg tccaggcagc      1380
tccagcgacc cttaaacctg accgcgcgca cgtccggccc gagggagcag aacaagaggc      1440
acccggaacc tcctccggcc agcaccacc ttcaccaggt tccgtcagtc gccaccact      1500
cccttcccgc gtccgcagcc ggcccagctg gggagcatgc gcagtggcgg gagccgggtt      1560
gcccgcgcca cagcaggtag ctgtactgca actgtcggcc caaaccaccc aatcaagaga      1620
cgtgttattg ccgccgaggt ggaactatgg caacggcgga ccaatcagaa ggcgcgttgt      1680
tgcccgagag cccctgccc cggcaggggg atgtggcgat ggggtgaggg catggggtgt      1740
gagcatccct gagccatcga tccgggaggg ccgcgggttc ccttgctttg ccgccgggag      1800
cggcgcacgc agccccgcac tcgcctaccc ggccccgggc ggccggcgcg cccatgcggc      1860
tggggcgga ggctgggagc ggggtggcgg cgccggcgcc cggggccggg cggtgattgg      1920
ccgcctgctg gccgcgactg agggccggga gggggcggg gagcgaggg ggagctcgct      1980
gcccgcgagc tgagaagatg ctgctgtccc tgggtgtcca cacgtactcc atgcgctacc      2040
tgctgcccag cgtcgtgctc ctgggcacgg cggccaccta cgtgttgccc tgggggggtct      2100
ggcggctgct ctccgccttc ctgcccggcc gcttctacca agcgtgggac gaccggctgt      2160

```

actgcgtcta	ccagagcatg	gtgctcttct	tcttcgagaa	ttacaccggg	gtccaggtga	2220
gccgcctccc	gctcccggtt	ctcgccgtcc	accgagctc	ccgggggcgc	ggacctctcc	2280
gctcccccac	agctggcgag	ggtcacccgg	ccggcccgcc	ggaccagca	cggagagcac	2340
gtgccgcctc	cccgccttcc	tctccgcatg	cttccctgcc	ttctgccgag	atcgctctct	2400
aggaagctgt	ggctgcgtcg	tcctgaggct	acgagtggga	cccgcgcgcc	ctttcccgcc	2460
ccctcgccctg	ggctctgatgc	tgcttagcaa	agtgggtgca	gatgcacgtt	ttaataata	2520
ggggcacgcgt	ttagcagttt	ctggcctttg	gtccaaagag	gtggctcatgt	tggaaacagat	2580
cggagacgtc	tacactccga	agtgcgcttt	tacagtgaac	tcttgaaaca	gaagtacaat	2640
tcggctctgt	gttctttccc	ctggacaagt	gaaagctggg	cgaagaaatg	aatacatttg	2700
ttaaccgtag	aagcctaact	agatacaatt	cttgccaact	ttaactgggc	ttgaatgtgt	2760
gggtgatctg	ttgtctgatt	actttctttc	tggtactgtt	tctctgtaga	gattggattc	2820
gtagattaaa	cttgagaaac	aaaccataaa	agtgggaagg	cctctttaac	agtaggtatt	2880
tgaagtgtta	taaaaaaaaa	aaaggtgaat	ttttctttta	tttctcagtt	tgaagaaca	2940
gctttattct	tggttattcc	taatgtccac	ctagtccctc	tttacttttc	ttggtagggt	3000
taggggtggca	tggggaaatg	ggacggatc	attttgctct	tttaactttt	ttttttcca	3060
cctacagcag	ctgtttttac	cctgtgggtc	gtcaggtaact	atatttagtt	tgcagttgca	3120
ctgctgatcg	acccttgatg	gcccagttg	gaagtgtgtt	ggggggaagg	aactaggaga	3180
ggccaggggcc	tccatttaaa	ccagtgctctg	taagtgtctc	cttgggaagg	aaaaagata	3240
ctgttccagg	tcatggtttc	ctggtagttg	acgtttaaaa	tgggcctcat	ttaaaaat	3300
caataattca	ggctaatttt	ttccctttat	atggttaact	caccaagttt	gtctaaatgt	3360
atgattttta	tcattgattaa	gtttttactt	ccacatcatg	tgacaactgg	cctgggtatg	3420
gatataagct	cagaacacaa	agtcattcac	ctgttaaaaa	aataattcta	tctgtggcgg	3480
gttatgttat	ttttgttcaa	agaggacaca	atatgatgca	gaatacacca	ttgaaggatt	3540
ttttggtttg	gcaagttctt	atttttttaa	atggctgtaa	aacctagcag	tgtttctgaa	3600
attgcatacc	ttacctgatg	ttcagagatc	cgatttactt	cttgatttcc	cagcaagtga	3660
ttttgaaaac	atttaattcta	atcattcccc	ccaccgtctg	ttcaaatcaa	aggaagtggc	3720
atccagcact	aattttcatg	catttatgaa	aggatgcctg	aggaccctta	agtataattc	3780
aaaattttgt	ttaatgtgtg	ttccttgatg	aagttcttta	ggagtctgtg	aacgaactga	3840
ttggcccatg	atcatcaaat	gcaagttagt	aacatttaat	aaaaatttaa	aaccaagagt	3900
ttcttggtcc	tgcattttta	tttttattgt	atggaggggg	caaatatta	ttttctgttt	3960
agtaacagag	cagggtattt	tgaatttatt	agggtctttt	tctgcagtct	gggtttcctg	4020
tgtaacaaa	gctacctttc	aatatttttt	attgtttctg	ttaagattaa	atcaatagag	4080
gaataaatag	ctatcttcaa	acataagacc	caaaggaaaa	agatttatag	tgatgttctg	4140
tcaccttatt	ttttacctgt	gactttgtac	cattaaactt	gtcactgaga	tgttttgatt	4200
aaaattttta	gcttgctttt	cttggtttgt	taggacactc	tttttttctt	gaattgtttt	4260
tatcagcttt	cgtttgcaag	gctagtgatg	attctcttgt	tctgtataaa	aatctgttga	4320
ctcattttctg	aagggagttt	tagtaattta	agagggtata	agtttttaaa	taaaagggtt	4380
attaattttat	atatattaaa	gaggcatttt	aaaataaaat	tttttttaaa	tgacattttt	4440
acacctttca	actctaggtt	taaaaaataa	gtgggtcaca	gtagttcttg	cagaagaata	4500
ttttctttta	catagaattt	ttaagctgaa	gagaagtagt	agtaggtcca	tgagattttat	4560
gatctgtgct	tggcaggtaa	acctgcttcc	aacaaattta	gttggtttt	tcttggtatt	4620
tgggtaataa	cctttttctt	ccccagtttc	actactttat	tttcatatgt	atctctgaga	4680
tagagaaata	tttcagtcag	tgctgctaaa	attgttccct	ataactcggt	tatcctttta	4740
gggtcctcca	gaatctctca	ttggtagtga	aactcaaatg	ggtactttct	tcaccattta	4800
tttctttaga	ataagtaata	agaattttat	aagctttttt	atatttcacg	taatttgaga	4860
ctattgaaaa	tccagttaag	tctctctact	gtgttgagag	gcattgatcc	aagtacctgt	4920
gttactttcc	tgtgctgcca	aaacagatca	cctcaaaact	agcggcttaa	aataatagaa	4980
cttaagtctt	cgtgattctg	gaggccagca	ctttgaaatc	aaggtgtagg	ctcaatttta	5040
ctccctctgg	aggccctagg	gggaatctgt	tcttggtggg	ttcaactctt	ggtgactggg	5100
ggcattcctt	ggcttggggc	cccatcactt	caacctctgc	cttacagtc	ttgctgccac	5160
ctcttctgtc	tcacatctca	ctctcccttt	ctcttagaag	gatgcttgct	attgggttta	5220
gagccacctt	ggatattccg	ggatgatctc	ttcatctcaa	gatccttaat	tataactgca	5280
aagagccttt	ttccaaataa	gaaaacattc	acaggttcca	gggttaggga	tggtggacaca	5340
ttttttgagg	ggctgccctt	cattccccca	caacaatgaa	ctccatagtt	ctgcctattc	5400
agtattttgt	agttatttctg	tagtttaact	tgcttatttt	ctttaggtat	ttacgtatta	5460
aagcattttg	gtctctgctt	tctttaacag	agaacctggg	tttctgtaat	aagtttactt	5520
actttcccat	aatcttttag	tttcttattt	acagatttac	cttcacatat	cccttaagta	5580
gaacatttga	ttaactgttt	tattttcgga	acaaactctg	attctgtata	ataaccaact	5640
tattcatatt	tcgggtattct	tttaattctt	atctgattct	gaaattacca	tcttggtgatt	5700
atatatatat	atatatggaa	ataactgaaa	tcttgataaa	ttaaaggtga	tataacttct	5760
aagacaatta	attatgtatg	atgtgggtgaa	tatactgggtg	tttgggttgt	ttgccactta	5820

aaagccctat	ctataggata	ggaagtaact	tgaatgtgga	atgcttagag	actcagagta	5880
agaggccgta	tatatatcct	tgagctggag	tttaaggaaa	acttatggga	aattaaaagg	5940
aaagttggag	tactgacaga	ggattgcgta	ggactcatga	aaaaggaatg	aagttacctt	6000
aaattctatc	atcgtgagtt	aacgtgaaac	tagatttatg	ttagtttata	gcctagaatt	6060
ctatccctagg	aatctagata	tatcctaaat	gttgagatag	ctgcataaac	aataactgta	6120
atcgttatga	taaataatga	caaactcttt	tagcatgttt	tgtgaagctg	ataaatgtta	6180
ataggatgtc	ttcaaatgtc	agaattcttt	tttctttgct	tcttttttaa	aaaatttctt	6240
ttcccccat	ccatgcaat	acactgaaaa	ctgatcattg	aaatttgtag	gccaaaaaat	6300
taatcaacac	gtaatagatt	ggggtttggg	tttttttgag	tcagggtcctt	cttctgtcac	6360
ccaggctctg	gtgcgggtggc	accatcatgg	ctcattgcag	ccttgaatgc	ctgggttcaa	6420
gtgatccctc	ggagtagctg	ccgtgccatt	atttctagct	aatttttaaa	agtttttcta	6480
gaaatggggt	ctttctgtgt	tgcccagggt	ggctttgaat	tccctggcctc	agggtatcct	6540
tctgccttgg	cctcccaaag	tgctgggatt	acagggtgtga	gccaccatgc	ctagcccccta	6600
ataaatattc	taattaccga	tttatcttgc	ttaaatcagt	tggtaacact	tgggaatttac	6660
ttcagaatat	attttacatt	agtggctctg	actgctaatt	cccccttctc	caaatgctaa	6720
tgtaatatata	caataaaatg	cacagttctt	aagtttatat	aaaataaaca	ggttttcagt	6780
tgacctgctt	taagtgtaaa	atagtggtga	aaacacaaga	aagaagataa	agaatttaag	6840
attttgacat	ttctctaata	tgcccttaac	ttctccaagg	attcatactt	ttttttgtaa	6900
gcagagaatct	cacactgttg	cccaaacacag	agggtgcagt	gtgcagtctc	cactcactgc	6960
aacctctgcc	cccgggtcca	agcgggtcctc	ccacctcagc	ctcctgagta	gctgggacta	7020
caggtagaca	gcaccatgcc	cagctaattt	tttttttttg	tatttttttag	tgggggtaga	7080
gacgagattt	tgccatattg	cccagctctg	ttttgagctc	ctggggtcaa	gtgatccgtc	7140
cttgatccac	catgcttagc	tgattcatac	tcttaactga	aacattgttc	caagtttctc	7200
agaaacagtc	aaggcttttt	atctagagaa	catttataac	tggtatcttc	tttgtgtagc	7260
actgattcat	caaactaatc	ctaaactcct	aatgagttaa	atttatattc	tgaatcttgc	7320
tgtaaaagca	gccatttcatt	agaatgaaac	atgtttactt	agaattggag	aagggagctt	7380
ataagtcatc	tagtctactc	ccttttatga	cacttctaca	ttctttctgc	acttctgcca	7440
aaatgttgcc	cagcgtctgc	tctgatacct	atagtcctaa	caagaatatg	aatcatacct	7500
gtatccctta	attttactct	tctctgctta	tttgccattc	atgtgaagac	cttaaataga	7560
tcttaaatgt	cttccctcac	tttagctgag	agtgcagga	ctgtgtaggt	gtgggtgtgt	7620
ttctgcattt	gcttatttaa	gcaggataat	aaaaactttt	actataggaa	attaaacatt	7680
tcccaatcaa	atacaattcc	agtctaacac	aattaaattc	tggttaggga	actgcttaac	7740
ttactagact	tataggaaaa	tactaaaaaa	atgtaactag	aactctattt	ttacacttta	7800
taaatataaaa	cctctgtgaa	caaaccagtt	atctcagggt	gcatttgtgt	atagtttttt	7860
aatgcctgat	ttttctattt	taaaatcaca	gatgcaatta	tacattcaaa	cactgccaca	7920
atacttttag	aaagttaaa	tttccctac	tcctacactg	cgtacacctt	tcctaggtag	7980
atcccagttt	gggtgtgtaac	tttagatttc	ttccaagagc	ttttgagtaa	gtgtttgaa	8040
tgtgggaagg	ttcttttagt	aaatgaactt	cttacagatc	agtttttttag	tacagtagca	8100
cgaaatatac	ctgcatacct	atggggatac	ctctgtgcca	ttacgatgga	aggcacggga	8160
aaacagcact	ccgtatatat	ctagtttact	ttccctcttt	tgtatatttg	tctgattttg	8220
tggagctgat	gcttctcaag	tggaatcaga	agttaaactt	tcctttacta	ttttctcatt	8280
ttattatgg	ttcttaacta	gaggttgatg	ttagtgtgtg	gaccattcaa	tagtaagtaa	8340
tgacttttca	gtaagggtat	tctagaacct	agatccctta	attcctgcaa	tattcccgtg	8400
tgtacattgt	tccagggtgt	gtcctgggta	ccaagggata	caatgtttga	tagacaatgt	8460
acctgccatt	atggagggtca	cattcttagt	tgggaagaca	aacaataaca	agaaaatgaa	8520
aatttactgt	gccatgccag	gttggttagc	ctgggtgggtg	agaggtaggg	gtttggaaaa	8580
tcttactgag	caagtgcacat	ttgtgtggag	ctctgtaaaa	gggccagctt	ggaaggtaat	8640
gtagtcatcc	agggtgagaaa	tgatgggttag	gggagtgga	agagtggtatg	ttaagattga	8700
aaagaattcc	aaatctattt	tagtggtagc	tgatagggtc	ttgtgattga	atgtggagga	8760
aaaagaagag	ggtgggttag	taacacactc	agtcgcagtt	agtgagtgct	gctgtgtgca	8820
agtattgttc	tattatgtaa	ataattccat	ctttacaaag	taggcaccat	tcttctctct	8880
ttacagacaa	ggaaaaggga	acacccatgg	ttcacatctg	tagtagccta	gccaggagtt	8940
tcaggcactt	attttctgaa	gatgctctgc	ctggcaatgt	ggttatattg	gttgaaatga	9000
gacccctac	tttcaaggta	ttcatctagg	aaagacatga	actgcccaatt	acaatatagg	9060
ataacactga	aattagagac	gtgtttatta	actttgccat	acagaggtaa	agtaactctt	9120
taaagtaact	ctttgcttgg	gttagtggtg	aaggctataa	aaattacttg	gagtttttac	9180
tttgaacatg	cgtaattaac	atggaaatgt	tagggaaaaag	agggtttcaa	ttgataacat	9240
aataaacatg	aggagtttga	agcatggcat	tcaagggttt	ctaaattctg	ccccgggttaa	9300
cttttccatt	cggttggttc	attctagtct	agcttttctt	tctgggcccgc	ccctcccccac	9360
attagaccgc	tctctctggt	aattccaact	caagcccttg	cttttctcca	tctgtcatga	9420
tgttacccca	tctcattgtc	agggttaact	ttatgtaata	ttaacatata	taatactgat	9480

ataacattag	catatttttaa	tgtatggatc	atctcctctg	caacattgta	acctcttggg	9540
gatggcaata	atgggaagaa	tgacttgatt	ttactttttc	ttttaacaaa	aatgggtggag	9600
tagtctgggc	acggtgtggc	tcatgcctgt	aatcccagca	ttttgggagg	ccaaggagg	9660
tggtacactt	gaggtcaggc	attcagagacc	agtctggcca	acattgtgaa	accccatctc	9720
tacaaaaaaa	atacaaacac	ttactgggca	tggtggtgtg	tgccctgtagt	cctagctact	9780
caggaggctg	aggtgggaga	atcacttgaa	catgggaggt	agaggctcca	gcttgggcga	9840
cagagtgaga	ccctgttcca	aaagaaaaaa	aaggtaaaag	ggccagggtgc	ggagggtcac	9900
gctggtaatc	caagcacttt	gggagggtga	ggcaatggat	cacctgaggt	cgggagttcg	9960
agatcagcct	gaccaacatg	gagaaacccc	ttctctacta	aaaatacaaa	attagccggg	10020
cggtggtgtg	cctgcctgta	atctaagcta	catgggaggg	tgaggcagga	gaatcacttg	10080
aaccaggag	acagagggtg	tggtgagcca	agatggcacc	attgcactcc	cgactgggca	10140
acaagagcga	aattccgtct	caaaaacaa	aaacaaacaa	aacaaacag	agagaaaagg	10200
cagagtactc	tagggaattc	tagtctgtgt	ttctgtggaa	atgtatatga	atctcacttt	10260
taagggatgg	agatttttga	atggcataac	tagttgataa	gttttgctct	aacagggtac	10320
ccaagtctag	tgagtccgat	tcattctttc	cttaaataga	tgaaggagga	agaacatcca	10380
ctccaccctc	aagagtaagg	cagaatgagc	aaagtacagag	aagttaaaaa	agaattctca	10440
cgagccagc	agtgcagaga	aaccttgggt	tagttgtgaa	tcaaaaccag	tactttttgt	10500
aatttttgag	cctatgcaat	tctccaaggt	tttatgttgt	ttcttctgtt	tctctgtagg	10560
caccagaaat	caaaacccca	aataagaaag	tgttacttga	agattttaga	gtacttattt	10620
gtgtataagt	gtaagtgata	tttggagac	gactttactg	cgctcctcca	gcttggcatg	10680
agaattccag	ggcggaag	aaaggagggt	gatggtacct	ggaaaggaga	gtcatgttaa	10740
gtcccagcca	catattaagt	gctaaccacc	tactgtttaa	aggtgtaagt	ttctagactg	10800
acaaaataca	tagtctctac	cgtaaaagtaa	cacataattt	agcagtgcag	aaagattgtca	10860
cttaaaagaa	aacttgaata	tatgtctgaga	tagttcacaa	attaaagaaa	tgaacaaaga	10920
actgagga	taaaggagga	atacaactgt	gtccaaatga	atacttaact	gggtgggagc	10980
tgttgcatat	gtaagcaggt	ggttcaccta	aaagtggat	gtaacgtagt	taacgccagc	11040
tcttgggtgca	cttacctatt	gcattgtctc	cgggcttaat	ttgtgttcat	ataggaataa	11100
attttttgtt	gggttttaat	tttactcctt	gtaattccgt	gggtgatatt	caaaagtga	11160
aaaattacat	aagctttctaa	tatatgagaa	gtcttctcac	ttgacatttt	ttatttggaa	11220
tttttgcaga	gagtagtttt	gtcacagtca	aaagattttg	ggatcttgca	gtgagaaacc	11280
taggtgtaat	tcctatttct	ctgccattcc	gtatgtcatc	tggttaagt	gtcaacttct	11340
cagtctcaag	attctcgtcc	ttaaatggaa	tactttttgt	catgctattt	tgaagacaaa	11400
atgagataat	acgtgaaact	gcctagctca	gtgaatggta	catcatagat	actcagaaaa	11460
aacacaccct	ctaaaataag	aacagtacca	aaagacagga	tgtaaaataa	gggcagtagc	11520
aaaagacaca	tgcatgctga	gtgtatgaga	aagaactttg	tgcccttctt	gggtggcaca	11580
ggccatggca	gttccacagc	atgacgtggg	tgctgtgggt	ggtagagcag	acatgccgct	11640
ccccgtcaact	gcctggcttt	gatgcttgct	ttcttcagct	gagaggacgc	agctgtgata	11700
tgaaggctct	gtgtgtacag	tcgtgacctc	acatttccaa	tttctgctg	gcagaaccca	11760
cagtctacaa	cgtacgagca	ccagagtgtg	cgtgagacag	acagcataca	gaggcttgta	11820
acatccttct	ggaaaacact	gtgtaagctt	tcagtgcgaa	taaacatgat	cagtggcaag	11880
ttctgttaga	tgtagtctgc	aagcatcctg	attttactgg	gcaagactat	gttgatttac	11940
aggcggtga	tgattccatg	gatagccac	tactagtatt	ttcacaatt	tcacaagaca	12000
ttcttactgg	aagattgccc	tggtcttatg	atactgctgc	ccttttagct	tcatttgctg	12060
ttcagactaa	acttggagac	tacagtcagt	cagagaactt	gctaggccac	ctctcaggtt	12120
attctttcat	tcctgatcat	cctcaaaatt	ttgaaaaaga	aattgtaaaa	attacatcag	12180
caacatatag	gcttatgtcc	ttgagaagca	gcagttaatt	acctaaccac	agcaagtagc	12240
ttagaactct	gtggagtgtg	attgcactat	gcaagggtatc	aagtaacaa	aaaattatga	12300
ttggaatgat	gtcaagagga	attctgattt	ataacaggct	atgaatgagt	acctttccat	12360
ggtcgaagat	tgtaaaaaatt	tggttttaagt	gcaaacagtt	ttttattcag	ctttgaaaa	12420
gacttgcata	aatctggaga	aagattatca	ggatttaata	tggtgaatta	tatggcatgt	12480
aaacatttgt	ggaaagcaag	tttagaacat	cacatattct	tctgttttga	cagaccactt	12540
ccaactagaa	agaatttttt	tgacatttat	tttacattag	gttcaaaatt	cctaattgcat	12600
gggtgggagaa	ctgaagtcca	gttagttcag	tatggcaaa	aaaaggcaaa	taaagacaga	12660
ctacttgcag	gatcctcaag	taagccattg	acgtggaaat	taatagtttg	ggaagttagta	12720
ggcaggaatt	caatatctga	tgaaaagatt	agaaacataa	agccttccat	cacaattccc	12780
acccggaaca	ggaattccta	ctcatcaaaa	ttctgcattc	atacaagagg	gaacctgatt	12840
atgaccatct	tctgttggtc	atttggtaga	ttatgtgggt	cacacttctt	ccaaatattt	12900
gcaaatcaga	catcaccatt	atcagcacia	gctaataagca	tcattcttga	atcatcacta	12960
ttacaggaca	cccctggaga	tggttagcct	ccagctttac	cacccaaaca	agctaagaaa	13020
aactgttggg	accaaattca	ttattttacat	tttcaacaag	atctggaaga	tcataatta	13080
gaaacgttga	tgttctatct	tctcttaaaa	aatctgctcc	taatggtggt	attctacatg	13140

ataatcgtgt	tctaataccga	gtgaacctga	cgaaaatgga	aggtttggag	tcaatgcaaa	13200
gggggatatg	atpagaagat	gtctgtgatc	gtgtcctgag	aagcaccagg	aacacctttg	13260
acctcagtg	ctctcgattg	aagagaagac	caagttgtat	tgatcagtg	ttgggaacttt	13320
acagaacaca	cccatgattg	ggttgctctg	ctttttaaag	ccaactgtga	gagacattct	13380
ggggaactca	tgcttctagt	tctacctatg	ctgcatatga	tgtagtgga	gaagtgtctag	13440
aaaatgagac	agacttccag	tacattctgg	agaaagcccc	actagatagt	gtccaccagg	13500
atgaccatgt	gctgtgggag	tcagtgatcc	agctaaccga	gggcttatcg	ctggaaacatt	13560
ctggacacaa	tttgcataac	ttatcaaaaa	aaaacttgga	atgacaattt	ctgggtgccag	13620
attaccttag	aacctttgca	aaaatagata	gagatagttt	tccttatgat	gttacctggc	13680
ttatttttaa	aggtaatgaa	aactacatca	gtgtaattcc	agcatcataa	gtcagaacag	13740
tgcttgtcaa	ggggcgcttac	cacacacttg	aacagatttt	tggcagatga	cttgggaaca	13800
aggctcctcc	atgtttgtaa	tggtgaccac	acaagttgaa	tggtggcagag	ttaaatgacc	13860
ccaatatggg	ccagaaccca	caggaagttc	atcctatgga	tgctaccaag	ccttctgccca	13920
ctgagaagaa	ggaagcactg	tctttatctt	caggaagatc	acactgctgt	ttaaccaaga	13980
gaaaaattag	agagtcacat	atcacgcaga	tccagtacag	agggtggcct	gacctgggag	14040
accctgatga	ttcagtgact	ttctggattt	tggttttcat	atgcaaaaata	agagggtctag	14100
caaggaaaaa	ccccttgttg	tttcttgag	tgctggagtt	ggaagaacca	gcgttcttaa	14160
tactatggaa	acagccatgt	gtctcattga	tctcattgaa	tgcatgcagc	cagtttattc	14220
actagacatg	gtaagaacaa	tgagagagca	gtgagccgtg	atggtccaaa	cacctagtca	14280
ttacagtttt	gcgtgtgaag	tactattttg	aaagcttatg	aagaaggctt	gcataaaaaa	14340
agcaaaaagga	aaaaaagAAC	tttgtcatct	gttaggttcc	atattattgca	tgataattgt	14400
gtttgtattg	attattgggc	aagtagctgt	ttgctatttt	gatcttattt	cagaagggca	14460
taataatttt	actattcaat	gaaacgtttt	aaacggggta	gaaaaagact	agtttttcta	14520
tgctttacag	cagaaatctt	ataatgatta	actggttaata	tatttcgttg	ggtgtgttat	14580
acatttaaaa	gttcaagtaa	ttataaacat	tgtaaatgtt	atatgtaatc	atattgaaat	14640
tgaaattctt	tatagctgta	cttctgtgta	atcaaagact	ggggagagat	agactagcta	14700
gctctttctc	ttatccatta	atcacttaac	agagttttga	ataaaaagtt	ccatttctatg	14760
ggataagaat	aatgacaggt	taacctattt	tagttgggta	ctatgttcta	ggtgtgttat	14820
gaagtagttt	acatagtttc	actgatttca	ctacaatccc	aggaggagta	gttactatta	14880
ttacactcat	tttacaggca	aagaaatagg	tttggagggg	ttgggtgttt	tgcccaagtt	14940
ctcatcgtaa	aatgacagat	gaggattcaa	attcaagtct	taattgaagt	ccattacttt	15000
agaacctacc	tcttagtggc	tcttatgtta	cagataaagg	gagagcagag	tgcttcttta	15060
cccttgtagg	gtagctaggg	cttgtgaatt	aagagactga	ttaacaggag	aagaggcata	15120
cacattttat	tgacgttagt	atttttacat	gcacagggaa	ggagggtttt	atttttattt	15180
ttatttttat	ctttattttta	aagagacagg	ggtcttgctg	tggtggccag	gctggactca	15240
aactcctgaa	gccaagcgat	tcttctgctt	gagattcctg	agtagcaggg	actatagggtg	15300
tgctcctctg	tgcttggcta	aagaaggggt	ttgtatgtga	tttttaacaa	aggctgataa	15360
attgtgaaga	agtgcactagt	caaaggagaa	gaggatttca	gctcccaggg	gtggtaaat	15420
gtgggaagat	gactaggaaa	tgtagatgaa	taaggtttgc	tatgcagggt	tattttgcca	15480
gtttctgggc	gcttaataag	ggacagggaa	acacctttac	agatggaaat	tcatatcacc	15540
tttccacagg	gaaatttatg	tctgcctta	ggcagttagg	ggaagggcag	agaattcttc	15600
ctgtatctgc	tgtgtctcag	gtgccttcag	ctcaaaaata	tccttatgcc	aaagtagcat	15660
atttgggtgt	ggcatattct	ctgatctctt	tcaacagcat	catctatact	taacaacagc	15720
aaaagttttt	tttaaaaaat	catgtttcaa	gatttgcag	tggaagacaa	atggacatga	15780
ttgagataaa	tgaagaatat	atatttttta	acaaagaatg	ctgtatat	atgtctctgt	15840
gacattgtgt	tatggaggct	aaggtgttaa	gcatgtgatt	acttttagatg	ccgtatgact	15900
acctgttttt	aagattaaaa	aagaatcaat	aggcagttta	tatgcattggg	agcaagttta	15960
aaacaacaca	gatgtgatga	aggcgagggtg	aaactggctc	gcatctaatt	caggccttct	16020
cctgaaagcc	agtgtgtgca	agataaataa	gtttgtttga	cgaaagcaga	ataactagtt	16080
tgctccttgt	gatgaagata	gttattcaga	aatcattttt	attggctacc	tctgaattaa	16140
taaatgaaaa	gagaaatttt	tttttctgta	ggggatgtct	gatgagttct	taaaaagtgg	16200
atgaacctga	aattatcatg	aacaagcaat	tataatgaac	ttaaaattac	ttaaagagtt	16260
atgaaaaaca	aaaagaaaag	ccgtatgttt	tcttgtgcct	tattttgaag	tgacaaatta	16320
tttgagggtg	acattttag	acggaactaa	tgtgatttaa	aaaatgagta	ctagatttac	16380
agaatgatgc	ctttaaaaag	tcactgggtg	actttaatta	ttttatttat	gtttattctg	16440
aaactacctt	tattttgaaa	atgaggtata	gctttgccta	ctgggtgacaa	aagtgtaaat	16500
aattcagtaa	acatctgtta	aaaaccagct	tggtgctagg	ctcttggggg	agaaaactga	16560
tcaggccatt	gaggagctca	tagtccctaa	ggggctgggg	acttgtcatt	agggtgtgac	16620
tgtgttctgg	atgctcctga	aggagtgtgg	gcagggtgcg	accaccatgc	ctggctaacc	16680
ttttataaat	tatgtagaga	cagggtctgg	ctgtgctgcc	catgctgggt	ttgaacttct	16740
gggcttaaga	gatcttccct	ccctgcccct	accgaccccg	cccggccact	ccacctcagc	16800

ctccccaaag	cactgggatt	gcaggcatgg	gccactatgc	ctgggctgtg	caaaactttt	16860
aaatcagtg	atactcaatg	gtcttgatgc	aattctggct	tggtggtaag	agaatgggga	16920
tttactcaca	agccâcagatg	tcactttttaa	ctctgaacag	atcaagctat	tggtattact	16980
catttatgtc	atcgataaac	tttatgaata	aaaactcatt	gtgcaaatat	ttaaacatac	17040
tacatacata	gcactgtgca	gtttctaagg	aaagtaatgg	aaacctttgt	cacatccctg	17100
gcttcagaa	ctttatgta	tctaagtga	ttgtctgca	aagttgtgg	gttaattgcc	17160
cccttctttc	ttctcttttt	aagatattaa	taaatagtgt	catgaccaa	agataatcct	17220
tatggacaag	atagatctaa	aaagccttag	ctaatttata	atcttgcata	atccatgatg	17280
acaagatgca	gaaacaaaa	tgcccagaat	aaaaacttag	caccattagc	agccatttcc	17340
ttttaagtct	ttacaagtat	actcccagtt	tcttgaaaa	tttattctaa	aatatgtaag	17400
acacacaaaa	cagcagaagg	actaatacag	gtacatcgaa	cacctgtgtg	cctaccgccc	17460
agtttaaaaa	taaactggaa	tgatgtttct	ctcatactta	cagaataaag	ttttaatctt	17520
tagcatggaa	ttcaaaagac	ttctgcccatt	ccagttcaga	gccacccttc	tggtctcctt	17580
gcttctcagc	cgcgacactg	cccatgtacc	caacaggcct	ccagggttac	tgcttccatt	17640
cgttcttatt	ctcatgaaca	ttttccttca	tctcatctgc	cagaatccta	cctaataata	17700
ctcctgctct	gcagtttaca	gttcttttaa	attaaaaaag	gttggtgacc	ctttagtgtc	17760
ctgaaaaaag	aaaaaacaaa	tttaaaacct	taaaaaggta	ccatattttc	atagtatttg	17820
cgttatgtct	cattacagtt	cctgtggaca	tgtctgtctc	ttttactaga	ttgattgtgg	17880
gctctttgaa	ggaagatata	tcttatgaac	agtgttttat	atattgttag	caatcaatga	17940
atgcttgcta	tatttttctc	atgaggatata	tgattattct	attttaattt	attaccgcta	18000
acctgtacta	tacataactg	ctttctgtac	ctgagctatt	tatgatctct	gaggtccttg	18060
tgagaaatct	aatttttgtt	aatcatggat	ggaaatattc	acaacatcat	tcgtcagttt	18120
cttcacattg	tcttcctttg	tattattacag	atgtttttaa	atatcaaaag	aatgtttttt	18180
tggttttatct	tttagatatt	gctatatgga	gatttgccaa	aaaataaaga	aaatataata	18240
tatttagcaa	atcatcaaag	cacaggtttg	tatttcattt	gcataaaaac	taggtttttc	18300
tacagatggc	acatgggcat	tcaaaatacc	gttcttatat	ttaaatgaag	tggtgttttt	18360
aaaacagcaa	ttttctgtgc	agatattaca	cctgttcttg	tatttttgtg	attttacttt	18420
ttggaaagtc	agaaacttga	aagctatgaa	ttttcctaaa	cttaccttct	ccctctgttg	18480
gatgtaagta	agctatcttc	ttacttgctt	gctttgtttt	tcctttgtgt	agctctttaa	18540
agagtgtatt	cattcttttt	gtaagtgatg	tttctagaag	tagcattggg	gggtcgaagt	18600
gtgtatacat	tttacatttt	tgattgtcaa	gctgcagaaa	agctgtattg	gtatgtaagt	18660
actcgtttcc	ttactatgct	cgctatttct	agtgtctgct	cttcttttcc	ttcttcaaat	18720
gggttttggtt	taattctagt	tgctactgtt	ccatcagagg	aattgcagag	aactgggtctt	18780
caaaacagtg	cagtatatat	tttaggtgaa	gatacttcta	aaaacctttg	tatttttgagg	18840
taattctaga	gtcccaagaa	tttgcaaaaa	gagtacattg	tcagcaaatat	ttttcccaat	18900
ggtgacatct	taataataact	gtagcacagt	agcagaatca	ggaaattgtc	attgggtaag	18960
gtacttttta	attctccaaa	taattcagcc	ctccaaaaaa	atcccacttc	ttatgttttc	19020
aaacctgtag	ctacttttga	tgctacttcc	ctaaattgca	tttttattac	tttaaaaaat	19080
ataataccta	gaagctcaaa	gctggaaaca	gcctgatcaa	tatagtactc	ttaagctaaa	19140
aacaacctga	tcaatatagt	actcttaggg	aaatcactta	tgccgtgtgg	ttttttttaa	19200
ttttcttctc	gtcagctgtc	tcttcatgat	tttgtgggtt	ttattactgc	ttataccata	19260
gatgagggtat	agaaagtaaa	agaagttaaa	atgcattttt	ctcaatttag	tgaatttaag	19320
attacattca	gatttatagg	acaagggttg	aagctacaag	gggttgatag	gaatcttgat	19380
gtatctgagt	attttcccca	actttattac	atgactgggt	cagactattt	tatctaatta	19440
catttctactc	ttggcagaaa	tagcaaaaca	gtcaaccaat	ggtcaatgct	gctgagaact	19500
ctggcctgtg	cagacatatt	ggctgtttta	cttctaatac	cattctgctt	ttcctgtctc	19560
gctgctgatg	gattgtttctt	ccagggtttta	aatacaaac	aaaagggatc	tggtggccca	19620
gtacagggaa	tggtctcttga	tagattttgat	tttcttgcat	ttcctttatt	ttgatccagt	19680
gttaattttca	tgtagagttg	tctgtttaac	aggattctct	taaaattcct	tcttcagttt	19740
acctgccagc	ttttctttgt	ccagggtttca	gtatgaactc	cactcgatta	atagagctct	19800
ctagtagtga	cttggtggagt	gggttctctg	aacatttctg	gaagtgttgc	tgatagtgtat	19860
aatattgatc	actagtactg	ttaattttgtg	tgcttactac	atgttggtt	ttatatgtat	19920
tccttcagat	taaggacttc	tagaaaacat	ccatgaaaaa	acagattaaa	aaaaacaatt	19980
ctgcatgtat	ttgggactag	aagggtactat	gggaaggata	atcttcatac	tcagaccata	20040
ctgacctgaa	tttcattttat	cagtttagag	aaccacttcc	ccttcccttc	accttacctc	20100
cgagtgcctg	tgacttttga	tcaccgctct	ggcaccacat	cctcatccca	gcaggatttg	20160
ggaaggctgc	tttttgaaag	ccttttataa	ttctgtaagt	tgagaaaaata	ctaggggaat	20220
gatttttaaat	ttcttttagaa	ttacaggctt	tagtcagtat	atgacagagc	cttttcttag	20280
aaaaatgtgc	atataaaaaat	ttgcatgtag	ttttagggtt	tcagagaccc	ctaaagccta	20340
tccatagacg	tggttcattg	tctgattgtg	tttaggtacc	cttctaaaaa	ccttttgaga	20400
tgtaggaat	cacaacagag	tattctctgaa	aatgtaatta	gcggaaagaa	catttcaaat	20460

actgtgtgttc	tgcttagact	ttctagtttg	tcttctgcca	ggcttgccgg	aataaatgag	20520
tttctgtggc	tgatactcaa	aagaattgac	atttaaatta	gtctctctct	tccttgtttt	20580
tcgcttgaca	catccttgtc	tctacattct	gtctctgtct	ctgttagctt	atctctctct	20640
cgagtcagca	ggatatagtg	gctgttattt	cttcccccta	tccttcaacg	atctactttt	20700
gacaacactt	tgcttttttt	tttttgagat	ggagttttcac	tcttgttgcc	caggctgggt	20760
gtaatggtgc	aatctcagct	cactgcaacc	tttgctctcc	gggttcaagc	cattttctctg	20820
cctcagcctc	ccgagtagct	gggattacag	acatgcacca	ccacgcctgg	ctaattttgt	20880
attttcagta	gagatggggt	ttcaccatgt	tggtcaggct	ggctttgaac	tcctgacctc	20940
aggtgatctg	cctgcctcgg	cctcccaaa	tgacgggatt	acaggcgtga	gccactgtgc	21000
cctgcctgct	atgtgccttt	ttaatctcat	gaaatgttct	cttttcttgg	ctgaagtgtc	21060
acttttcttg	ttgaacagca	tgctgtgtga	gtagaatgtt	ataaaaagg	atggactttg	21120
gagtttagaga	gacccagggt	cctgttcggc	attgcagaaa	tgctgttctg	caataggctg	21180
tgtgtcagtg	ggcaaatatc	ttatctctca	gagccttatt	ggtaagggtg	gagtgatagc	21240
tcctttcagg	caccttacag	aggctgtctc	ctaactcctg	tagcgtacct	ggctcataga	21300
tggcatttaa	aagtgggtgt	gatgacagtc	atagctcacc	attagcatag	cgctggatcc	21360
atggcaggga	agcgcgtcac	atgcagtatc	tcttggacta	cacagggccc	tcctgaatta	21420
ggactgctg	tttcatgagg	atagggatga	ggaaattaga	cttgcctgcc	ctcactgcct	21480
tcactctctc	tcctccaagt	taatgggaac	tatgactctg	ctttggcttg	attgccatgg	21540
aagattctca	cacagccaaa	tttattgcta	tcttagttaa	attatgccag	aacacaaaat	21600
atgaagtatt	tgtcaaagta	atataatctc	agctgtaact	gagatagtca	gaaactgtct	21660
gtaattctgat	gtcctatctg	aaaggtagct	gagaataaac	aagaaataaa	gagaattcag	21720
tagcaaatat	tggtagacac	aagcttttat	attttgacta	gttaagctag	ttcttaaatg	21780
tttccactaa	aatattcaag	tttaaggcca	tagcccagg	cagcttatta	tgaacatgat	21840
gtattttgga	aatctttcac	tttctcttaa	aagttcttgg	gaggggcatg	tgaggccata	21900
atataaccat	aaaaccattt	gttttaaaat	aaaaccattt	tttaaaattc	ttccaaaata	21960
aaaaattatt	gcaggaaaaa	atgctaacc	tggtttttaa	ctttgtacgc	caactatatt	22020
tccaagatgt	gctgtagcct	ggtaaccata	cagaaccata	cagaattagt	tctcagaatt	22080
tattgtctgc	ttacttttgc	atttggtaca	ggataaacag	ggctgattat	atgggttcta	22140
agacatgact	agaaagaaat	atgtttatca	gttattattt	cttccatcta	aattagaagg	22200
ggctagggag	agggtctcaa	cagggaattta	tatactttag	agaaaagtga	tcattgatag	22260
cccaatagta	tagatatctc	aacccaataa	cacagggtgt	gtctgtctct	gggatcatac	22320
actgtagggg	agaatctttg	caagcaacat	tctacttata	gggagccata	acaaaagttt	22380
catatgtata	ataattataa	gtcttaagtc	atcaagaaaa	agttaacttg	tgaatgataa	22440
tcctcgatta	aaaagagaga	tgtataataa	tggataagag	atttttcttg	gttaattttt	22500
agtattaaaa	tggctaaatc	ttctttggga	tattctgact	agtatggtgc	attgtctaat	22560
agatttccca	tagctgagag	ctaactatct	tgtaatctgt	ggaaaactgt	cctctttggc	22620
taaaacttta	ttgttaattc	tctaaatcct	cagcttttat	tttctacaga	cttttttttt	22680
tttttaacat	ttbcttctc	tgactcactc	cttttgttct	cattttcatg	gcctgagaac	22740
atgggtgatg	atagaattat	tcttttcaca	gattaacagt	tttctttctg	agtatcgttg	22800
agctcatgtg	tgtattaact	agagaagtct	cccttacatt	tcatttttat	gttttcttct	22860
tcacccaggag	atagtttgta	gccatttact	ttcaaatacc	agtttctgct	gttcttaaga	22920
cctgtatcat	ttgtctcctg	aatttcactt	catttctctc	ttaaacctag	tcctctgttt	22980
cccactctct	gcacccactt	tgccacttcc	tgtttgttta	attggcaagg	gccactctct	23040
gtgttgga	ttttttcttt	ttgaaagctc	aactaacaac	ttctaggaag	ttttttattg	23100
ctactgttat	caattcatac	catcttacct	ttgtttttgc	aaccttttgt	taataacata	23160
tttatttaac	tatagttatt	agcagtctga	gatcatttta	cttggttaca	taaggagcac	23220
atatactctac	ccagcatcat	tgttaaggcat	gtgagacctt	tgtttgattg	ctgtcttaac	23280
ctagtaccga	gtcctaataa	ctcatttagta	gaagatgaag	tgctcttgcc	ttttgtctgaa	23340
catatatata	cacactgaat	atttagtggc	aattcatagt	tgcatattgg	cattttttgt	23400
ttataaatttc	ccctttctca	ttaaaaaac	tttgttttct	agactttagg	atttagagaa	23460
gtcatttttg	ttccatacac	atgctgctgt	tggattattt	aggatttttg	tgactgtatt	23520
ttatctttga	aataaaaagc	ctttcaagaa	atgcaaaaaa	aaaaagctca	aaaaacagaa	23580
aatgtatatt	ttttaaatat	ctcagataga	tttaaaagaa	ttttaaacat	cctaatacata	23640
gtacttttga	agcccatcca	tagtacaacc	tgtgaagagc	ctcatgtacg	cgctaactgg	23700
gtcctgtctc	tgagtttgac	tggattgttg	ctgacatctt	ggccatcagg	cagaatgcgc	23760
taggacatgt	gcgctacgtg	ctgaaagaag	ggtaaaaatg	gctgccattg	tatgggtgtt	23820
actttgctca	ggtaacttgt	ttccatgctt	ttctctctat	atatgtagtt	tataaaattt	23880
tttttttttt	tttgagacac	gtctcacttt	attgctcagg	ctgagtgacg	tgggtggaac	23940
acagctcact	gcagccttga	cctctggggc	tcaagtgaac	ctcctgcctc	tgctctccaa	24000
gtagttggga	ccgtagtgc	caccatcatg	cccggtctaa	ttttctat	tttgtagaga	24060
tgggggtctc	gctgtgttgc	ccaggctggt	cttggactca	agcaatctgc	ctgtctcagc	24120

ctacccaaat	gctggattat	agggtgtgaac	tgccataccc	aaccctataa	aaatgttata	24180
ttttaaaaat	taacaatata	cttcatgtga	atgtatgggt	tttaaaatgg	gttttaatagt	24240
ttattctcag	ttgaagtaat	tttgtttggc	attttttagtg	gtgtgtattt	atatacgtct	24300
gattatccat	atgcggtttt	ccttcagcat	ctgtggggat	tggttttaga	accaccacag	24360
ataccaaaat	ctaaggtgtt	caagaccctc	atatagaatg	ggatagtagt	tgcatataac	24420
ctgtgcacta	ctttaaatca	tctctagatt	acttataata	tctaatacat	tataaatgcc	24480
atgtaaaatg	ttgttatact	ttatttttta	tttgtattat	tttaattggt	atattatttt	24540
taatttttat	ttgttcacat	atttttgatc	tgtgatttgt	tgaatctgca	gatgtggaac	24600
tcattggatgt	gaagggccag	ctgcagtaaa	atgaaagagc	aaaaatgcaa	atgtacaaaag	24660
ttcaaacaaa	taggaaattt	aaaggcatag	aatttgatag	gcaattacat	taaactgttg	24720
ataacagtaa	ttagtgtatc	gtatgatatt	aaaaaaaaaa	agcaaatgtg	atatataaaa	24780
cttactttct	ccagtctctg	aggctagaca	tccaagatca	agggtgttgac	aggggttagtt	24840
tctcccaagg	cctctctccc	aggcttgtag	acagcatcct	tcttctctgt	tcctcagggtg	24900
gtttttttcc	ctgtgcccac	gcaccctctg	cactgcttcc	tcttcttaga	aggactagtt	24960
acactggatg	actaatcctt	ctacagagac	tgctaagggtc	ccactctgtg	gccctttttt	25020
aaccttaatt	accacctcta	agtcctctct	tctgaataca	gtcacagtgg	gaactattag	25080
ggcttttagta	gactgatttg	ggggaacaca	cttctgtccg	taacagtggc	acataaatat	25140
ctttagcagg	attgattttt	taaaatccct	aaagatcgtg	agtattgaca	tggttaaggac	25200
gcttttttagt	gactctgtaa	taagtgggtg	gaagaattgg	gagttaaatc	catctgatgg	25260
atcagggtttt	ttatttttta	aaatgtgtat	ttaagaaaga	aagcattttc	attttaactg	25320
ccaacaaaac	taaacttcat	gtgttttcca	atacagtgtc	acatgcagtt	tttttgaatt	25380
atgttgagac	aaggcaattt	tcagctaaat	gttcttttaga	agctaattgt	tgaagatatt	25440
aaatatagat	taaattctga	aatgtagttt	tcattctgta	ctttttgcaa	gagaagttgc	25500
cttttttgatg	actctggcca	attgttattt	taaaagtaaa	tgctctttct	cccgattttg	25560
ttgtggcagc	atggaggaat	ctatgtaag	cgcagtgtcca	aatttaacga	gaaagagatg	25620
cgaacaagat	tgcaagagta	cgtggacgca	ggaactccag	taagagccta	cccgttttta	25680
tttttcttac	cagctctcag	tttctaatt	taagaattaa	attaaaaatc	aagaattggt	25740
ttgacaatgt	attttcccat	gtgtaattac	taattcaggg	ttatgctgag	gtaacagaaa	25800
ccctctatgt	acaggtagcg	aggtttttca	gccatcagaa	agattgctgt	aaacaactag	25860
gtcctttgct	ggtcagtggg	ccttaaagag	gaataaaaag	agcatttggg	gtcgttcaga	25920
gtctataaat	agaactaact	gcattttaac	ctgacattta	agctagttaa	caagctcatc	25980
ttacttcttg	tcttctttag	tatcagattt	ggtttttagaa	gcagcaactg	ttttctgtaa	26040
ttgcaaattt	tgaatgtctt	acatgtacag	aaaaaccaaa	aaaggatgaa	tctctacaaa	26100
tggttaaatca	ttcagtgtaa	ataatatttt	ataaaacttt	attccacaaa	agtgggggaga	26160
gttcaatctg	ctttgtatag	aatgctgatt	gctgccaag	gcttttcccc	tggttcccc	26220
cggagacaaa	gcaccatgat	caccggggcg	acttgggctt	tctctttcag	tacatgacat	26280
gtgctcagaa	gcttagctcg	tggtgcacag	ctttcccttt	cctttctggc	tcctccctc	26340
tgtcttccct	cctctcctct	tgccctcccc	tcaccagggg	tcctgggcag	cagctggagc	26400
tcattggtgaa	ggaagaattc	ttcatgggtc	gctggcgag	tgccctgggt	gagcattggt	26460
tattcacatg	cctcttctag	gtgtttttac	attagaacat	tgcatctggt	ttgggcatgt	26520
gttgggtgac	agaagcagaa	tggaatgaga	tgaacagtga	ccctttatcc	tggttatagc	26580
aacccttgag	aaccaagctt	gggtgtctca	aagggtctgt	ttagtctgaa	acagtgtggt	26640
gaattttggg	agaattgtgg	tcattgcatg	taggtctcca	aaagacagaa	taagttggta	26700
atatgtgtta	tcgacttttt	acaaaaaaaa	tttaaaaatc	atgaatttat	accttaaaat	26760
gtccatccca	cttctctccc	agctgtccag	tcacccagc	aatggatgac	tgctgtggag	26820
ttccttctgt	gtcctgctgt	gggcattgta	tatatgaagc	aaatgaagat	agctgccttt	26880
tgggtgatgt	tgccatccta	tgccacagtg	tcccttgctt	ttttgcccc	atgaatatag	26940
ctgccagtgg	cgctagggct	gaaaaaatca	gctctttaca	cttgtcatgt	gtcctgttta	27000
tgtggctgcc	ttcgtgagtt	tcttcttggt	tttggtttgc	agcagtttaa	gtatcatata	27060
tctgagtgtc	atttaaaaat	ttttacctgg	attgggtctc	tgagcttggg	tctatgattt	27120
gggtgtctgt	attaattttg	gaaattttct	tgctcttatt	tccttaata	ttattcctac	27180
cccagtcctt	cttctccagt	tatgtttgtg	ttgggttcatt	tctcgtgtgt	ctttagttct	27240
tagatgcatt	attcgttttt	tggttggttt	tttttaaat	ttttttttta	cgccccctcc	27300
cttttttctt	tttgtgttac	attttggata	atttctgttg	accacacctt	gagttcatgg	27360
attcttccct	tggtgtgtgt	gagtcactag	gtgagccagt	ttaaggcact	cttcatctct	27420
gctactgcgt	gtttcattcc	tcacatttcc	ctttgaccct	gtttcatagt	ttccatctct	27480
gtgctagtgt	atctatctga	tcataaagct	tagtcacggt	ttccagttag	acctttatca	27540
ttttattata	cttgacgttc	tcttaaatcc	cctgtttgat	aattccaaca	ctcggggcat	27600
atctgagtct	gcaaattttg	attactttat	ctcttcagat	tgtgtcttat	cttgcccttg	27660
tcatacttcc	taagattttg	cctaaccgtg	ggcctttttt	gtaagacagg	agaaatggag	27720
gcaagttgtc	ttgatacctg	gaaatggata	gacttgctct	tctgcttggc	tttttagtgt	27780

gaggagtggga	gtcagtcac	tgaggaggtg	ctactgcattt	gggttttgct	catgtgcttt	27840
ttctcacagc	ttcaggtttc	tgtagaactc	attactttgt	ttgtaggttg	gggatgtcct	27900
cccgc tagag	cttttcctca	gtgtctat	cacactcagc	gttttcacat	agcaccttgg	27960
agtggctctc	ttctttatgc	ctttccccc	tatacttctt	ggatacttgt	tactgaactc	28020
tcgctagttt	ggtggtagaa	ggagagggaa	gggaagtgtc	ttttcattct	tagggagaat	28080
ctcaggggtg	gagccttctc	tgatcctgcc	ttgcttctgg	ctgtaagtct	gtgcccagta	28140
tgtattcctg	cctttactaa	gagtttttcc	ctgttctctt	cacccagcct	catcgagtat	28200
tcaccctg	cccatgggta	gcagggtttt	gttgcccctg	ttcatcagtt	tcagggtgct	28260
gttccatagg	aaaggtagaa	agaaggatgt	gggctgggccc	ctgagccctt	cccacagggc	28320
tgcttttccc	tcccacaagc	ctacatccag	tcttccctga	ccgcagtggtg	ttttcttttt	28380
tctttgtctt	gtgagtacac	aggaggtctg	tgggtcgagc	ctgtgaaatg	tgctgcattc	28440
tccttggtgc	tgtagcccag	gggttcgtct	gttccactgg	ctcatacttg	gctttctgca	28500
aaattgataa	aatTTTTagc	taaattcttt	ttactgggtat	ctgttacatt	ggcccccaac	28560
taaaacaacca	cttgcatctt	gtttctcctt	tgagttttcc	atctttcctt	agacttttgg	28620
gtaggttggt	tgctttgcaa	ccttgtagct	ctctgaaggg	tctaagaaaa	gtcatgaatc	28680
tacagcttgt	cagtgttggt	gtttgtgtag	gtttggcagt	agtattcctt	cagcattcta	28740
catacttaat	ggaagccgcc	tcccatTTTT	ggttaataaa	tttcaaaact	tggaacaatg	28800
ttagattttac	aaaaacgtca	gaaagaacag	agtgttccctg	tttattcttt	atatagcttt	28860
tttttttttt	tttttttttg	agttggagtc	tcgggtctgtc	acccaggtcg	gagtgtagtg	28920
gcacgactctt	ggctcactgc	aacctctgcc	tcacgggttc	aagcaatctt	ctgcctcagc	28980
ctcctgagta	gctgggatta	caggcgtgca	ccgccatgcc	cggctaattt	ttgtattttt	29040
agtagagaca	gggtttcacc	atgttggtcca	ggctgggtctc	gaactcctga	cctcttgatc	29100
cgccccgcctc	ggccccccac	agtgtctggga	ttatagggtg	gagccaccac	gcccagcctt	29160
cttcatctag	ctttaacatc	taattgtgac	atcttacata	acatgggtata	tatttctcaa	29220
aactaagaaa	taaacattgg	taccacacta	ttaatgttac	tacagatttt	tattcagact	29280
ttaccagggtt	ttccactaat	gtcctttttc	tgttctaaaa	tacaatccag	aatagataca	29340
aatccattca	acttcagtg	tttaaattat	tgtttttcat	tatatgaagt	gctgtgtggt	29400
ttttgtcaaa	tctgtttatt	tgggtttta	cttcaagctt	gtctttgttt	ctttaagtga	29460
taaaggcata	atttaaaagg	tgtgttgggt	tatttcagtg	cctaaagtct	tgtctgagtc	29520
acttggtttc	tgctgttctt	gcttatggta	ctttctttcc	ttgtttgctt	tgttatcttc	29580
ctttgtctgt	ggctgtgttt	ggttaagtta	tttgtggaaa	tcagtgaag	cctcaggtg	29640
gagtgtcttt	ctccggagaa	catttctacc	gttttttagct	gggcccctta	aggctcctt	29700
agcgtgggcc	ccacccaaac	gagattctga	gttgaagggtg	aactgagcca	ttcaggcagt	29760
gcagccagggt	ttgcagatgc	acgtgagacc	tgctcacctc	tcatttactt	tcacctgag	29820
agtagagcct	ttggtgtttc	gttcacttgt	ctgattctct	cttcacagtt	ctattagaag	29880
gtccatgggt	tttggtttct	gtgcccttca	cttatgagt	cttgtaaat	aaagttctgt	29940
tttatgctta	cttctgcttt	actgtgtttg	cttaatttca	gtcttaacat	cttgccaact	30000
cttggttact	tttaaaataa	tgttatatcc	agctttttta	gttggttttca	gtaggaagg	30060
tgattcaaat	aacctagtct	ggttatgggc	tacgagaata	gcctccctgt	tttttgtggg	30120
caaaattcca	gccttttatg	ttcctagcgc	agtggtgata	acagactggc	aggttcaaga	30180
ggcgtgctg	agcagctttc	actgtaagg	cactgtccca	ggctgggttt	ctaagaatct	30240
ggatgggtgt	ttcatttctt	aatatgtacg	ccctgtgaga	gcggtacat	cttgctcagg	30300
ttcttatgat	tcttttgttt	ctgaagggtga	attaagtaag	tgacatggta	gaatatgtta	30360
agtcacacttt	cgtgtggctt	actagttctc	atgaatctat	tccatgattg	tatcagttct	30420
tattcagtat	tagtatttaa	gaaatgcaga	attttgtttc	aaaaaatata	tttgtattat	30480
aagttgtgaa	gaaatacatc	tccataatta	ttgtcgggac	aatacagtat	tttcttaagg	30540
aacttattgg	ttgtggatgc	aaatgaagca	tatttgtgat	aaaaataact	aatagaagtc	30600
attttgttag	actatgagct	agtaaaaact	atggcacaaa	catggagact	taacactttt	30660
tcttccagct	ttcacttaag	ttccttttca	gataggaggc	agcctggtgg	ataagagtat	30720
tgggttttgaa	attagattca	ggtttaaatc	ccagatcttc	tgtttaattc	ttattttatt	30780
tcaggtagat	tttctggata	acttgctata	gcttatagct	cagtaactgc	cacttcaatt	30840
ttatgttatg	gagagacggc	ttctttcctt	aaacctcacg	aaccaacctc	tgctagcttc	30900
taagtttttt	cctgccactt	ctttacctct	ctcagccttc	agagaattaa	agggagttag	30960
ggccttgctc	tggattagga	tttgctttta	gggagtgttg	tggctggttt	gatgtttttt	31020
ctagagcact	caaactttct	ccatatcagc	aataaggctg	ttttgctttc	taatcattca	31080
tgtgttccagt	gaagtagcac	ttttaattct	ctttaagaac	ttttcctttg	catccgcaac	31140
ttggctgttt	agtggaaagg	acctagcttt	tgacctacct	tggctttcaa	cataccttcc	31200
tcactaagcc	atttctagct	attgatgtaa	agtgaagagc	atgcaactct	tcctttcact	31260
ggaacgctta	gcagccattg	tagggttatt	aattggccta	atttcaatat	tgtgtgtctt	31320
caggggaatag	ggaaacccaa	ggggcggtag	agagaaagag	agacaggaga	acagggccatc	31380
attggagcag	tcagaacaca	cacgacattt	atcaattaaa	tttgtcatct	tatatgggtg	31440

caattcatgg	cacccccaaa	caattacaat	agtaacatca	gagatcacag	atcacataaa	31500
cagatataat	aatatgaaat	attgtgagat	taccgaaata	tgacacagag	acgtgaggtg	31560
agcacatact	gttggaaaaa	tggcaccaat	agacttgctc	gatgcagggt	tgctataaac	31620
cttcaatggg	aaaaaaatgc	aatttccgtg	aagctcagta	aagcgaagca	tgataaaatg	31680
agatgagcct	gtcactccta	agaatgttcc	tgtacaagtt	ttttgcatct	gttacttacc	31740
ttttcctatt	tgtgaatagt	atcttttttg	agtagctgtg	tttttttatt	ttatatacatt	31800
tatatgtatc	ttttgaagaa	catactttta	agcttaattt	attgattttt	tttctctcat	31860
aatttccact	ttttgtatcc	tatttaagaa	gtccttgcca	aacttaaggt	tgctaagatt	31920
ttctcctttg	ttttcttctg	gaaattttag	agttttgctt	ttacatttag	ttctaggatt	31980
tatttataat	taatgttttc	atatgggtga	agatcggaagt	tcataatttt	ttaatatagg	32040
taaccatcac	tatagaaaag	attattttccc	cccaatgttt	gaaataagta	gactgaatat	32100
agatgggtct	gttatcccta	gatcaatgga	gcatttgctc	tggttatattg	atctatatat	32160
atatatcctt	atgccaatac	catactgtct	taataatgct	tgctttgcag	taagttttta	32220
aatagtgtag	ttgtcttcta	aatttgttct	ttcttttcaa	agttgttttg	ttctatttag	32280
gttttttgca	tttctgtgtg	aattatagaa	ttagctcgac	aatttctacc	caaagtttgt	32340
gggcttttca	ttttgattgt	attgaagata	tagatgaatt	tggaagaagt	tgatataaca	32400
ggattgaatc	tttggattca	tgaacgtagc	ctgcatttgt	ttacttaggt	cttctttatt	32460
tatttcagtg	tgttttgtg	tttaatgtac	agatttgcac	atcttttgcc	agatatatacc	32520
ctaagaattt	cagtttttga	tactattgta	gatgacattt	aaaaaaattt	caagtttttg	32580
tttgttgacc	taggcatata	tttgactttt	taataacta	accttgctaa	acttatttat	32640
catctagtaa	cttacaaaat	atattcctta	ggatttctta	cataaaacat	catgtcattg	32700
tttttagaat	aacagtttta	ctttgtcctt	tttaactctg	atggctttta	tttctttttc	32760
ttgctaaaat	ttctggctag	acctcctagt	acagccttga	ctagaactgg	tgtaggggaa	32820
atcctttcca	tattcctcat	ctttagggaa	aagcactcat	tcttttatcc	attcttttag	32880
tcctagcccc	attgcccttc	ctaaaatttt	tctctcatt	ttccttcac	acaccttggt	32940
ctttttcttt	gcaatcatat	catgatatgt	aacgacatgt	ttttatttat	ctgtttaatg	33000
tatttctttt	cctcacttgt	ccatgaaggg	aaggaccata	tgtgttgtta	tcctttgtgc	33060
agttcctgga	acataataag	tatataagaa	atagtttctg	aattagctgt	gaatgaattc	33120
atgccttctt	gctgtctgtc	aatgttcttt	taaattaaac	atctaagaca	gcaataata	33180
ccacatgagt	tatttaacctg	agaaataatc	gttttattta	taaatgactg	agttgaaagc	33240
tgatagccca	cagtaattgc	tttcatggct	ttgaataata	accttactgt	tacaaaacac	33300
attttcatga	aaatgaatgt	gtggtgtttg	gaactagctt	taatgtttgt	cttctgtttt	33360
ttccttctag	ttgctataat	ataataagga	attttgtatg	tttttcttaa	ttgtaccac	33420
ttttctacat	tttcttaaca	gatctgggtga	atcttcatta	ttaaatataa	ttatacatat	33480
aaattattgt	ttaataataa	tattaattat	taaaaataat	ataaattatt	aaatataaag	33540
atacatataa	tattatctgt	taatttctaa	gttaggtgtg	ggttctgaag	actattatat	33600
gaatgaacaa	aaagcttgca	tatttgctgt	gaagctgaaa	gtacgaaatt	tttagatacc	33660
attataccag	tatctaaaga	aaaaattcag	taccacatag	gtttttaagt	aggagctgta	33720
tgatcatagg	tcattccagat	gaagggaaggc	ttctgtacca	gacgtacaga	ggtagacagt	33780
gttggtctgag	tactgtctga	gatctggcaa	gaatgaatcc	aataaacgta	gttttctccc	33840
atgagctcct	gtcttgtttc	ctgtattctg	tttgtatttg	aaaagatttg	gtgtgcataa	33900
cttatttttg	tcttttggct	gtcaatcaaa	gttatttagt	tagtttttgt	aactcagttc	33960
tcaagctagg	agtttttgct	gtataatttt	aatgtttctg	tttttacttt	cctaagcaga	34020
taagcgtaaa	aacttagact	aattgattac	ttattaaacg	tccagcttga	tattcttctt	34080
tatattattt	tagtttcagt	ttatataaca	aatgaggttt	cttataaata	aaatttaaaa	34140
tgactaaaag	gagctgtgtg	aaatagggaat	tctgtgtgaa	gcttttgaat	gtgaacattt	34200
agaacgtttc	acatgggtggg	aatttactat	atgattttca	tcaaatgagg	tacttttttag	34260
tggttggtact	taacgatact	gatttctaaa	atttgtattt	ctaaaaatga	cgtattacag	34320
gatctgaaag	ggcaaaaact	cattgaggtc	ttgtatgagt	cagcgtttca	tgccctattt	34380
ttaatttagt	aattatttagc	atataattag	aaatgttttt	agattcttca	tggtgacct	34440
accaatgaat	gtagcactgc	atttaaaata	tagttcacgt	tatgttcata	cttaattggt	34500
gcatttttgt	tgccctctct	gaaacgaagg	tcacatgtaa	ataaatatac	attttctctt	34560
actgtaggaa	atactctgtt	agcatttagta	ggttagctt	ttttagggtta	acaataacaa	34620
aaacaaagct	cacacaaaat	aaaccaaatt	tgctctatgt	cccacagatg	tatcttgtga	34680
tttttccaga	aggtacaagg	tataatccag	agcaaacaaa	agtcctttca	gctagtccagg	34740
catttgctgc	ccaacgtggt	aagtaaaaaat	ttgagtgttt	gaacaaataa	ttttcaaaga	34800
taataacatt	tttagttttt	cttctggaa	aagatacttt	tggtttacag	ttgaagggaat	34860
gaatgtattc	attccttgaa	ttagtgtaca	tattatctct	taggaaatga	agtttcttct	34920
ccttaattca	ctttcatgct	attattacat	atatctgaga	aattaaagttg	aagtgtctgt	34980
tacgatacat	attcttgtgc	catggattta	tttaaaatct	atctaagtac	atgattatgt	35040
agatggaagc	tttttctaca	gtgtatgggt	tatatgtaat	ggagcttctg	ttttgtaaga	35100

tgacagacct	aagttggagt	ccaaactcgt	acttttatta	gctgtatggt	tgcaacttgg	35160
aagttgtgta	atggttgctga	gcttgcttct	tcactcttta	aaagaacata	tgccctataa	35220
gtagatctaa	atctgtgtga	ggattagatt	agaaaatatg	tcaagtttct	attgggagaag	35280
ttacacaaaag	ttgggtccaca	gtgcttggaa	gctgttaatg	tcttcaacaa	tggtaatgtt	35340
cttaatatcc	atatttttaga	aaattgaata	attgggtacac	caataagcta	tgcaatttaa	35400
ccaaattggg	aagtatacag	aaaacagtgg	ctatgctatg	ttcttagagg	tgcttttgaa	35460
gcttgactgt	gatttagtgt	gtgatctcca	tatgttgata	gtcactcact	gagcaaatat	35520
cttggtgggtg	acattacagc	agggcctatg	acagtgtgtg	ctaattggaac	tttctgcaat	35580
aatgggtaaa	ttcttcatct	gttctgtcca	gtgtgtctgg	tcctaccaat	gtgggttttg	35640
agcattcaac	atgtgactag	tgcatgaaac	taatttttaa	ttttatttaa	tttttagttta	35700
attaaaaata	agggggaggt	tttacaaggt	gcttacaaga	gcagatatgt	cataggtata	35760
tgacatcatt	tgtaacagta	cttttaaaaa	atgcccagttt	gttttttaaac	acatgtccta	35820
ttaaagtaagg	agtgtctcag	aataggagggt	ttcagttgggt	ctcccatctt	gccagctctc	35880
ttttgacttt	cattgtcttc	ctctgtctaat	agacatgacg	ttctgtctat	tcagttgtct	35940
ttttgcaatg	ccattgtctc	ttttgccctt	ttcacattta	ttaaacagaa	caaaacaaaa	36000
accactctcg	aatctgtagt	ctacccttgt	tgtaagcact	ttttccagta	ctcactctgc	36060
cctcaatttg	ttttgggtctg	atttgaaatt	ctctccctag	acttctgtgg	ggctgttctc	36120
cattatcttc	ccaactctct	ggcgattact	tcctagcctc	ctttccagcc	ctcttctgct	36180
tcatttcttc	ctgtacatg	tgttatttcc	agtgtcaggt	tttgggtgtt	gattaatctt	36240
actttttgtt	tctcatggtg	gccttctctt	aaatccatgg	ctttagccat	cgtttctctg	36300
actgctgatg	actcgcaaaa	gcttccctcc	ctccatgtct	ctctgcctaa	ctctggacc	36360
attgttacia	ttgtccatta	gagagcttgc	cttgactggc	ccaaaaggat	gtctcaaat	36420
cagcatattg	aagatagaat	ttatccttcc	atgcatacac	tcataattct	tgtcttggtg	36480
actccatcat	tcagtttttt	tgccctaagtt	ttattcacaa	aaagaacaaa	ttgatagcag	36540
ttgcatacct	cttataggaa	acttagacat	ggaggaagaa	gctgttcaga	tggggtcctg	36600
cagaagtga	ggcactgtgg	taatatttaa	acttttctca	gctgttcgaa	gggtttgtt	36660
ttactaatt	ttccttagac	ttgttttagg	tatttggtt	tctaattggt	ataagggatg	36720
tggaattaaa	tgtatcttaa	tctgccacct	ggaccattta	aagtaagccc	ctatggtggt	36780
tttttttttt	aattgccatg	gttaaaacca	tagttgctag	cgaaggtgac	atacttaagc	36840
tttttgaaact	ctcttaaaaag	aaaacagaaa	tttaatgatg	tgtctataat	ggcaaacacg	36900
atacctagaa	tttccatggt	attcataggg	tgaataacac	tggtgatgtg	agagatttga	36960
gagttctttc	aaaacaggag	aacaaaggga	ataagctaca	aagcaatttt	tttctttgta	37020
gacttaactg	aataaaaaatt	atttttatgt	ctcaaacatc	atatgaacaa	atttagttgg	37080
caaatggcaa	gctaataata	ttttataata	taggatatta	atatacttaa	tattacaaaa	37140
gtgcttcata	attagaaaag	acataaacta	gaaaaatggg	aaaagggtcat	gaataagaaa	37200
ttcaagagat	acaaatgacc	cacacacttg	aacaaatggt	tattctttct	cataatcaaa	37260
gaagtagaaa	ttaaatgaat	actttgaagc	caacttctga	gaaagcatag	caaacaagaa	37320
agctagtgtc	cagctttgtg	tggtaacggc	actctcgtct	ttaaagaagg	gtgtttgtct	37380
cctgtggctg	ctctcaggca	gggccacaaa	cttggtggct	taaaacacca	cagatttctt	37440
ctcttacatt	tgagaagtct	gaaatgggtc	ttactcagct	gaaatcaagg	tggtggcagg	37500
gctgcagtc	tttgtggagg	cttgggggga	tttgttctc	ctgtacgggg	tcctgtgctt	37560
gggttcgggg	cctgtgcttg	gtctgggata	ctgtgcttgg	ttcgaggtcc	tgtgtgggt	37620
ccagtgtctc	gcttttacca	ccttgaagtt	catctggaaa	tggtcactgg	tcgcccacac	37680
catatagctg	actctgggtt	tccctctctc	tcactcgtct	taaacctgtg	tttttggtg	37740
atttctaate	tctcttttct	tggtccctct	gcagcttgca	gggccttctg	cagctcttgt	37800
ctgccccagc	ccccgggtct	gcccattcca	gtgctgggt	gttctgttcc	tgccctgctt	37860
ttcctcagcc	cttggcaacc	ctgtttgttt	tctcccttcc	ttagcagttg	agaacatcgt	37920
aagatcaatg	ctgactgctc	tctgcagcca	agccaggcca	tttcatttca	gccgagccaa	37980
gtctgtgtgg	agcagttctt	ttatttttct	ccttttgact	acctcatggt	tttcacggat	38040
ttttgttctc	ttcacattca	aggatttttt	gctttcagaa	agttatattt	ctctggaaag	38100
agtgacacca	atatcccttt	tgatttcaaa	atcttaatgt	ggagtctctt	gacttggatt	38160
tttttggaag	aaactgctga	agctgccatg	tctaagaaga	aaacttttga	gaaaaatttt	38220
cttcttagac	atggcaacgt	caacagtttc	taagctcttg	attccgtcta	ccctgtctcc	38280
atcggtgcct	cagtcactcg	ccttacttct	ctgcagggtt	ttctccagc	ttgcaaatgt	38340
actccaattc	tgaaataact	aagtctatag	ctgtgcaaag	agaagtctgg	gccccttgct	38400
ttcttgtgtt	tgactccatc	cactctccag	aaatgaatcc	cacttctcac	tttaaccactg	38460
acctccaaag	catcgatatc	tttgtgtcag	ttgtcatatt	tgtaactttt	cacataactt	38520
ttgacattat	ttataccttt	ataaccagga	aataatttta	actttattgt	agaaataaac	38580
aatggagtat	aatttttctt	gttgaagata	aatatcacct	cctcttctct	taaacatctc	38640
ttccctttgt	ttttgtatta	cattgggttt	cccccttttt	ttatttctctg	ggttgtcgta	38700
ttccctgtta	ttattttttac	cttttttttt	ttaatgtgga	tggttccgga	gtctgtattt	38760

cttgccctttt	catcttctgc	cctttattat	tctcagccac	tgccattact	tcagttatcc	38820
attcccatgg	tttccacatg	cttagcttcg	gttgattctt	gccattttac	agaccatatt	38880
tccaactact	tctagaatgt	tttggtcctt	cagcctcagt	atgcccaatt	tgaactcatg	38940
ttctctctcc	cccttcttcc	ttccttcttt	ctttcgtctt	ctctcccttc	cttcttttct	39000
ttccctccct	cccttcttcc	cttccctcac	tcgttctctc	ttgcttgctt	gcttctctct	39060
ctctctctct	tttcttcttg	ctttcttctt	attcttctcc	ctccctctct	tccttctctc	39120
ccccactccc	caacttccag	gctaaagcag	tcctcctgag	tagttaggac	tacagacata	39180
cacgtgccac	cgcgccccag	tcctgtgtct	ctttgtttcc	ctgctcctcg	ctcttccact	39240
tatcttttga	tggcaggtgg	gtgcacgcag	gcatgctctg	catgtcttcc	tcttggccat	39300
tccccctcta	gttatgggtg	ggctttatct	acgcgttctg	gagcagaagc	ctagtccaaa	39360
agctattttt	ttaaaacatt	catgataatt	catttccctt	tatgttttaa	aaatactagc	39420
tttctgtctt	tatttcttta	ctaacttact	tggatgccag	taattagttg	ttttagtgaa	39480
caccacagag	tgatattttg	aaactttgga	cttcataaag	ttggatgagc	tccagtagca	39540
aagaagggaag	tgtaacttag	tttaactgac	aaataaatgc	ttccagctt	gggtgtgcgat	39600
tgagattttt	gttgcaagtt	tgtgaatcaa	tttaactgcc	cctgcccctg	ggactaaagt	39660
cagatcacgtg	cttggtggaa	tctttgtctt	tcccacacca	ccctgcattt	taaaacctct	39720
tgtgtgggac	agtcccacca	tgaatatgct	gttcttctct	actcagctac	tttccctcca	39780
gagaggccag	tagaaaaatc	agactagtct	tttatagtct	attttcatgt	cacttattga	39840
gagctactgt	tttctgttaa	attgtcagta	aatattttta	tcaaggaaaa	gggaggcaat	39900
aggaaggaga	gaagaacaaa	tccttaaccc	tagtaggaac	ctaataatga	ggatttgttc	39960
tggaataattg	cagtagtccc	ccagctaaa	aaccttttaa	aaatatgtca	gatataacca	40020
agaggattga	aatcgatatg	tcatacaaaa	gcttgttcac	ctgcagcctt	catatgcaat	40080
tcctatgaat	gttcatagca	gcattattca	taatagccaa	agtatggatg	caaccctaat	40140
gtccatgaag	caattaatag	gtaaacaaaa	tgtgatctgt	tcacacagtg	gaataactac	40200
tattcagcca	taaaaaggaa	tgaagcactg	agtcctgcag	ccacacagat	gaacctcaga	40260
tcctatgctga	gcgaaggaa	ccagaaacag	gaggccatgt	gctgtgtgac	tgtatttcta	40320
ggaaatcttg	agtcaccatg	ggcaagatgc	tatcaccttt	gttcagtggc	cagaagcgag	40380
ggcactaata	tttacccttg	ccggggtcta	ctagattgaa	gcgtttccgc	tagggcataa	40440
acttccaaca	cggtgacttg	tacatgtaga	tatttgatca	atatatagca	aatgaatatt	40500
gattttaaaca	gaaaaaggca	agtgaagtg	ctttctaaac	ttagagccct	aaatatatga	40560
ggttgtggaa	ttatagatt	ctgttgtgtg	tgtttgaggg	aatttaaaaa	taatttagat	40620
gttaaacagt	atattgtgga	gggtttttgt	aactaattaa	tgacggcact	gaattgactt	40680
ctaggccctg	cagtattaaa	acatgtgcta	acaccacgaa	taaaggcaac	tcacgttgct	40740
tttgattgca	tgaagaatta	tttagatgca	atttatgatg	ttacggtggt	ttatgaaggg	40800
aaagacgatg	gagggcagcg	aagagagtca	ccgaccatga	cgggtaagtg	tgttcacgca	40860
cctgaaatgc	ctgtacacgg	tatatacagt	gcacatgttt	atgtagaatt	cagttttaca	40920
aagttaggta	agtgactttt	tttctccat	tacatttacc	cggatatatt	ttcaagatgt	40980
tattaagatg	taacagtgga	gatttcatta	gtcctgcata	gtgtggtatt	tcttggctgt	41040
cgtgtgagtc	ctgtggactc	accaattatc	attaatccag	cctctttcta	ctcaaagttc	41100
acacttaaaa	ggaaaagctc	gtaaaaggga	ggaagacgtg	aagaaggagc	acgcctggca	41160
gtactgagtg	cacgttatta	gtcagtgtcg	cccttttgct	gtatttttctg	taaaatattt	41220
attaaatttg	ggtgtcattg	tgacaagaag	aaatgcagtt	aagtgtgacc	tttttttttc	41280
cccaaacatg	ttaggtttta	agaacctttg	agctattgtc	agatataacc	agaaaaaat	41340
agaattttta	gtgagcagga	taacttagtt	aaactaacca	aacatagttg	tagctgttag	41400
agaaatgtaa	acatggaaat	aggcaaacag	ggaagtgtgt	ggagtctctg	tttcttttct	41460
aaaatatctg	tttgagctgg	gggtgagaga	gaacactagg	cttcatgggg	ttttttgttt	41520
tttcgttttt	tgttttgaga	caagagtttc	gctctgtcgc	ccaggctgga	gtgcagtggc	41580
gcaatcttgg	ctcactgcaa	cctccgcctc	ccacgttcac	acgattctcc	tgcccttagcc	41640
tcctgagtag	ctggaactac	atgcgtgtgc	caccatgcac	gactaatatt	tgtattttta	41700
gtagatatgg	gatttcacct	tgttggccag	gctggtctca	aactccttac	ctcaggtgat	41760
ccacgcacat	cggcctccca	aatgagcttt	gtgtttttac	ctcatcagct	gtttgggggt	41820
gagccactat	gtatgtcagt	gtgcttgtat	cagtaggac	tactgagggc	agatgttcaa	41880
aatatgagcc	tccagcacgt	tttacatgga	aaccttcacc	tgaagcattc	gtctgaagtt	41940
gatgtgcctt	ggaaatttta	tagagtaata	tttttaacta	caacaaaaca	tttataaaaag	42000
tagacattat	taaagcattc	agaagtgagc	aaggatagaa	attattctgc	ccaaccttac	42060
acgtaggcct	tctagacgta	gtactgtgca	cogttacatt	atctaact	gtctgtgtgt	42120
catcttttga	tgttagggat	ttttccaaag	ttcagtgaga	ttatagttgt	caaagtatta	42180
gtctgttaaa	taagtataag	atgagggtca	ctcaggtttt	aaaagaaaag	ctctttgact	42240
gaaagagaga	gcagctgtct	actgcagaaa	gttagggagg	gaggctggag	gagtgaggcc	42300
caggggctag	ctagtataaa	aattgggtat	ggctcgaagga	aaaaaaaatg	taacatattt	42360
atatctgaaa	gatgattgtt	ctcataattg	tatataaacac	agagtaattg	taaagtagaa	42420

aactaagggtg	tttttcattt	tagatgtaaa	tgtttagaat	atgtaatgca	tcagttttaa	42480
aattaaaact	gtacgaaatg	cacagtgaag	cgtcttcctt	gctttccacc	ctgctacctg	42540
gccttccctt	ctccttccta	gcgataacca	gttttcttaa	tttgtgtg	gttgatgtg	42600
caaatttaag	tatatcttct	tattctacca	tccctccctt	cttacagaaa	agtggcata	42660
taataattttt	ctctttttaa	ctatcgaagg	agttacttac	ctatttttgc	atttcaaaac	42720
agacagttca	tcaagattgt	cgttggttta	ttaaacaatg	tttaagatta	aacaagtgtt	42780
tataaccaat	gaaaaacaga	tagactcccc	ataataacct	tgtttaaatg	ctgctacttt	42840
tatcatgtcc	cctcctgtct	aagaaccctt	tgggttcagca	gagctcatgg	gtaaggccag	42900
cctctgttgc	ctgccatcgg	aggaatgcgt	tccagccgtg	atctctgcct	tgcccttcgt	42960
tctcctgtg	ctgtgccgtg	aagcctcggc	cgtgggtgaag	ctggctgact	gagtcctcct	43020
gcaccccatg	catattcagt	agttgaaggc	tttgtgtggc	caatcctgct	ttccacagga	43080
aaccaccctc	tcttttgttg	ccctcatcca	aggctactgt	tctcccagag	tgacaggcgg	43140
cacctttccc	agcatagcac	tgtgccttct	cctgcccctg	ctcttgcatg	actgctgttg	43200
cactgatggc	gtgtgttaca	gtgctggcac	ttagcacagg	gctctgcctt	tctctcttcc	43260
cagccgcata	ataagtgcct	tgagggaagg	aaaaccttct	gtgagttgca	ttgcctgggt	43320
tccaacctcc	cactgccctg	cttatcctct	gctacatgtg	agctgactgt	ggctttgggg	43380
tggctcactgc	ctatgtgtat	tcattacaaa	ttgtctcctt	ttgaaagatt	gacctttctg	43440
acttaccag	ataccataaa	gaaaataaaa	tcttatcact	tcagtcagg	ataaagtatt	43500
tctgaattaa	aggaaaaata	caccagagta	aaatcaagac	tgaaagacaa	actgggaaat	43560
tatttgcaac	ctagatcata	gaaaaggggt	catttccttc	ttgcgtaaag	tgcaattaca	43620
aattgataag	aagatgactg	ataactagaa	agaaaaatgg	gtaaagaaca	acaatagaca	43680
tttcacattt	aacctcattc	atgataaggt	aagtgcaaat	gaaaactaca	ggggatacct	43740
tttttttttt	ttaatccatt	agattggcaa	acatcccaag	gtttgatcat	aggctcagtg	43800
ggtgagattt	aagtattatc	aggcattttt	atactttgct	gttaggaatg	caatgtagta	43860
caaacctttg	tagaagttgc	tttggaaatg	tctctcagat	gtacaaatgc	attcacattt	43920
tagatttagc	attcccgctt	tctgagacat	tattcaacat	gtatacgtgt	gcacataaga	43980
tataataata	acacgttttt	ccttctagtg	tggtgctttt	aacctgtagc	ttgaaaaaac	44040
tctgctttca	ttgttttttt	ttgttttctg	tcactggctc	agccctgctt	tcaattgttt	44100
ataggaattg	atgggtgttc	tggtctgtgt	ataatctact	ttagttttaag	agtcacttta	44160
aatttatatga	catctgatat	aagttgtgtt	aggtagaaaa	ttctgttaact	tggaatactg	44220
taagtacttt	gtggccacat	ttcattagta	ttaaatatta	tctctatata	tagtaggcta	44280
tttaatatcc	atattttatg	atgcaattaa	gaaataattt	ttttctgaag	ttggtagatt	44340
gttgatatgc	catggcccag	tgtttctcaa	agcattctgg	gggatcagaa	tttgtcagaa	44400
ttagctgcag	tgattgttga	acatgcaggg	cctctgctcc	actccacgtt	gctaccagga	44460
cgctctgcag	gtgagagctg	ggaagctgta	gaagctgcag	tgctaacaaa	tgctacagga	44520
attcttgtag	tcaccttcat	gaggtcttat	gttgaggaga	ggcagccagt	agtgctccct	44580
gtccttcccg	ttttatgggt	taagtttcat	tttaaggagg	gtataaatca	aagccacact	44640
gggcattctc	tcattggttca	ctgcttcttg	taatcatgga	agatgtcatt	gcggcagaga	44700
cgaaacagtg	tagtttgatt	actattgatt	tttttttaat	tatttttctg	aagtggctgt	44760
tgtaatgtaa	ttaattgtgt	gcttaaggac	aacctttggt	attctatttg	agtattgtgt	44820
atgatcctag	ttaagttttt	tctaccagta	ttttcatatt	acaacataat	tactttccat	44880
ttctattaat	atttttatat	ttaaagtatg	gaggccgggc	acagtggctc	acgcgtgtaa	44940
tcccagcatt	ttgggatgct	gaggccgggtg	gatcacaagg	tcaggagtcc	tagaccagcg	45000
tgaccaacac	ggtgaaatcc	catctctact	aaaaatacaa	aaattagccg	ggcacagtgg	45060
taggcacctg	taattccagc	tactcaggag	gctgaggtag	gagaatcact	tgaatccggg	45120
aggcagcagt	tgcatgtgag	taagatcgtg	ccactggact	ctagcctggc	tgacagagca	45180
agaatccgcc	taaaaaaaa	gggatcaggg	aagaggggat	tacagataac	ccaaagaaga	45240
aggaaaaatc	tccacaagtt	cacctgtcca	gcggtaaccc	caatttgga	attttccttt	45300
aacaatttgg	atattttcct	tttaattctc	ttttttataa	tgtctatatg	ttggagagag	45360
tatgtgcctt	tapgtatttt	ttaaagatga	gatttctgtg	tgtgtctata	tctcctgttc	45420
ttcatatttt	cttgtgtgtt	ataaacagct	gtacatgtca	gtatatatac	ttccgtaact	45480
tttttttaaa	ggctatatag	tgttcattga	tgtgatttaa	cagcagttat	ctccccggct	45540
tcatcttgtt	ggaatgtggg	tctgtgtgtg	tgcttccaga	gcaaatgggg	cttgggtttt	45600
cagcaagtag	acctgtgacc	tgtacgaata	gttggaagac	tttctctatt	acccaagcgt	45660
atcagtatac	tttagtgctt	actagaaatt	tatgggtaga	aaaacaataa	tatcttagag	45720
tattttttcc	tagattccct	aagggtgctat	aggggtgattt	ttactcatgt	aacatgaact	45780
atgcttcaac	taagatagtt	tttgcaaatg	tggatatata	agtactttat	taaacctata	45840
ggaagtattt	ataccactta	tttctccctt	tcagtgttag	aacctcctaa	atggcatttg	45900
acattgaact	gctttccact	ttgtcgcatt	ctcctctcat	tgctccctac	tggtctctga	45960
accttaggga	cttggctgtt	atagcccccac	catgggtacg	ctgggcccctg	gtcgtctctg	46020
agacttagtt	tcttcatctt	acaaggagat	aataacagcc	cctgcctgcg	tagaattgca	46080

gagatcaaat	gaaataatta	acatactcaa	aagcatgccg	taaacacatt	ctgagcacat	46140
gtacgtttta	ggaaaaacaa	aaggacccat	gcacatttcg	gagtgccttt	gtctcagcag	46200
cactgcctct	tcttccaaag	ctgacgtctt	agtagaggcc	ctgccacgtc	ctgagcactg	46260
tactccacga	agcattctat	ttctgacatt	cgaaatgcag	tctgttccat	cttccttaca	46320
atctgtatgc	cagcacttga	aataccgggt	atctgcagtg	ttgaccaggt	gattacttaa	46380
ttatggaaat	gttgagggtg	agatctagat	aattcagtga	aggcaggaaa	attggtgtcg	46440
gaatctgtct	ttttatgtgt	cagaaataga	aataagatag	ggtgagaagt	aatttgtggc	46500
taaaacacta	taatagctaa	cacatagtgc	atactgtgtg	ccaagcactc	ctgtagggtg	46560
ttgaaatctt	ctattattat	tatccctact	ttatagactt	gcacccttag	gcacagagag	46620
gcggacagtt	gtccaagggt	accccagagg	tggagatcca	ggctacctga	ctccaccatg	46680
tgtgtctctc	cctagggcac	agttgtgtcg	ctaaaaatac	tttttaagca	gttctttgat	46740
tattcagatg	atagtactgt	agggaaaatta	agacaaaaat	aatgaaaaat	taaaaatctt	46800
attttagtgt	tttgacatg	tattattaaa	gccagtttac	tcctggaagt	gtgtaagaat	46860
acagggtatt	tttgatcacc	taaatgtctg	atgttactaa	gagctcgaca	ctgaagtcaa	46920
gaagagcagt	tgacagaggt	acttagcaaa	aacgggaagt	gtgtgggggt	gaaggagcaa	46980
agacaagtct	tcctcggacg	gtggagtgtg	gaattcatca	tttctcagaa	cacgtctttg	47040
aacgcatttt	caatttgagg	caaagggtct	cagcctccca	ctcggcatac	ctccctacct	47100
tagtcagctc	ttaaatctta	ggaatatttc	ttgttcttct	aaggaaacta	aatatgttaa	47160
cattcttacc	tgtccacagg	gagcccccta	caaagaaggg	agtttctagt	ctccgttctt	47220
tcttggaata	aataatagcc	tcataccttg	tgcaatcgag	gctgaaaaag	actgtctcct	47280
tttttcaaat	aagcaagtct	tagaaactac	agttgtttac	agggctcatg	gctattccac	47340
agtaataaatt	ttgggtcttt	taccaattat	ataatatgtt	aaaatatggc	aagtatcagg	47400
aaagcaagga	gtggcaatga	ttagaaacca	atggccaagt	tagagaggag	gggcaattgc	47460
tcccccaagt	ttgttgtggc	tgtgtagcag	tcagtgcaga	gaagctgtgt	gtcaggcgag	47520
aagcaaaagt	gaggattatc	agggcgctgt	gagtgccag	ctgtgtgcca	ggtcaggagg	47580
tgccatcggt	agccagacca	gcttctctct	ggccccgtgt	gagctcgcag	tctggtgggg	47640
aggcagcagt	caccatgggt	acaggtgaca	cactaggatg	gggctgggtg	tggtaggcag	47700
ttgcgggtcc	cttcagagag	gtgagtatgg	acttagagga	ggctccagct	tcctattcct	47760
gggtgtgtct	tagcactaaa	agttgtcaca	tgaaaaataa	catttggtag	tattgattta	47820
acttaatgac	ttatgtaatt	gtagttgact	tagaaattat	aacatgctct	tctacttcag	47880
cttgaaaacc	ccaaccacca	gtttataatc	cttttttttt	aacttttgg	tatttttctt	47940
aaggaattct	tactttttct	tcattttaca	actttttttg	tcctgttacc	ttattttcat	48000
ttttacttta	tatgacatg	agttctaaaa	tagtaaaaaa	aaagaattat	ttttgttctt	48060
tgttagaatt	tctctgcaaa	gaatgtccaa	aaattcatat	tcacattgat	cgtatcgaca	48120
aaaaagatgt	cccagaagaa	caagaacata	tgagaagatg	gctgcatgaa	cgtttccgaa	48180
tcaaagataa	gtgagtaaca	acagttccag	cacttccgga	acttcgggtc	aactagattt	48240
cagtatagtc	aacaatttga	aaccaatgta	aatgggttata	ttgtctcaag	aatacatttt	48300
ataaattcaa	atcaaatttt	atgcattgtc	gatcgtgttt	ttaaactttac	ttgtacaaat	48360
cagtctaaaa	gaacttggtt	cagtggggcc	atctacttgc	attgatagta	tttcttggac	48420
aatactacgt	gataacatag	caaattaaat	taaaaacaac	aacaaacaca	caaaaaaact	48480
ttccagtgtc	agatgcccg	acctacctgt	caggtcacat	aaagtgggtg	tactgtgtga	48540
ggctctggct	ttggggcag	gtgcgcagaa	aagcaaggga	ggggtagagg	actatgcgga	48600
cgtgcagggt	gacatgatgc	tgttatattt	gttgaaaata	gaagggggca	gttgacagcg	48660
ttatatccaa	agtgtcttct	gtgggttaatt	atattcagaa	attttagcca	attgttttat	48720
tctctaaaata	tgtactttct	gtcgaagaaa	ctatcattgt	tcttcttttc	cttgttttac	48780
agtacagtgt	ttttaattaa	ccctcctggg	ttacttttac	caggtgaaaa	tgattaaaaag	48840
tgtaataggt	taacaatgaa	actttaagct	tctatttttc	attgactctt	aactgtacat	48900
gatgtaatgt	attcagcgag	ccattcagga	ccactttggc	ccatggaaga	aatttaaaag	48960
taagatctac	atgtattgac	atgaaaatat	gttctcagaa	aaaagactaa	tgtatttaat	49020
gtcctactta	ttttataagt	atttagaata	cctctggaca	ttttaaaaca	atgattattg	49080
ctagggtgtg	tgattttata	agcaatagaa	gcgctttccc	tttctgtttg	tgttttagat	49140
tatttatatc	ggtagtttct	gctatcataa	ctttacaaat	cttatgtaat	atgggaaaaat	49200
gagtttaact	tgtgtgtttc	cttctttttc	ctgcctttct	aattctgtgg	gaataaaggc	49260
gtttttgaga	cagcccaggt	gtagtgcaga	gtccatatcc	atggattcca	cattcatgga	49320
ttccaccaag	cacagaccaa	aaatactcag	aaaaaaagg	ggctggctgt	gggtggtcat	49380
gcatgtaatc	ccagcacttt	gggaggctaa	ggcaggcaaa	ttgcttgagc	ccagaagttc	49440
aagacagcct	gggcaacatg	gtctctacag	aaaatacaaa	aattagccag	ctgagcctgg	49500
gcgtgcacct	gtagtcccag	ctactcagga	ggccgagggt	cgaggatcac	agagtgcctt	49560
aagggttgaga	ctgcagtgtg	ctatcattgt	gccaaactcca	gcctggtaac	attcaaatat	49620
ttttcaaaaa	aaaaaaaaaa	aaaggatttg	ggaggatatg	catatgttat	attcaaatat	49680
atgccatttt	attcatatat	cagggaactg	agcatccttt	gatcttgggt	tctgccgggt	49740

atcctgggac	cagccccctg	tcgatacaga	gggaccgctg	tctaagaacc	gctgggtccta	49800
tctttgactt	ctggcggaat	aggagctcca	tgtaaaaaag	aggagaagct	gcagcggggt	49860
attagccatt	tgtgagtcag	gtcactgtaa	aactttatca	aaagtttaaa	agacaaaaag	49920
catcctcata	aaatgcctta	aaaccacctg	ttgaaatatt	acatatataa	ttcatgtata	49980
ctaatacatag	agcatattaa	agatatttta	gaagactaga	aacttctatt	aaaccaagtt	50040
tctggatggt	tccgtattca	tccttatttt	ccagggacct	gcataaactt	tccagcgtgt	50100
aatagctacc	tgattgatat	tttttgaatt	gaaatactga	agtgactaaa	atctaaactt	50160
tttccattct	ggccatagga	tgcttataga	attttatgag	tcaccagatc	cagaaagaag	50220
aaaaagattt	cctgggaaaa	gtgttaattc	caaattaagt	atcaagaaga	ctttaccatc	50280
aatgttgatc	ttaagtgggt	tgactgcagg	catgcttatg	accgatgctg	gaaggaagct	50340
gtatgtgaac	acctggatat	atggaaccct	acttggctgc	ctgtgggtta	ctattaaagc	50400
atagacaagt	agctgtctcc	agacagtggt	atgtgctaca	ttgtctattt	ttggcggctg	50460
cacatgacat	caaattgttt	cctgaattta	ttaaggagtg	taataaaagc	ctgtgtgatt	50520
gaagattgga	taatagaatt	tgtgacgaaa	gctgatatgc	aatggtcttg	ggcaaacata	50580
cctgggttga	caacttttag	atcggggctg	ctggaagggt	aaaagctaaa	tggagtttct	50640
cctgctctgt	ccattttccta	tgaactaatg	acaacttgag	aaggctggga	ggatttgtga	50700
ttttgcaagt	cagatggctg	cattttttgag	cattaatttg	cagcgtattt	cactttttct	50760
gttattttca	atttattaca	acttgacagc	tccaagctct	tattactaaa	ctattttagta	50820
tcttgcagct	agttaaatatt	tcattctttg	cttattttcta	caagtcagtg	aaataaattg	50880
tatttaggaa	gtgtcaggat	gttcaaagga	aagggttaaaa	agtggttcag	gggaaaaagc	50940
ctgttttagc	acatgatttt	attgtattgc	gttattagct	gattttactc	attttattt	51000
tgcataaata	atttctaata	tttattgaaa	ttgtctaatt	tgacacacct	gtacacacag	51060
aaaatgggat	aaaatatgag	aacgaagttt	aaaattgtga	ctctgattca	ttatagcaga	51120
actttaaatt	tcccagcttt	ttgaagattt	aagctacgct	attagtactt	ccctttgtct	51180
gtgccataag	tgcttgaaaa	cgtaagggtt	ttctgttttg	ttttgttttt	ttaatatcaa	51240
aagagtcggt	gtgaaccttg	gttggacccc	aagttcacaa	gatttttaag	gtgatgagag	51300
cctgcagaca	ttctgcctag	atttactagc	gtgtgccttt	tgccctgctc	tctttgattt	51360
cacagaatat	tcattcagaa	gtcgcgtttc	tgtagtgtgg	tggattccca	ctgggctctg	51420
gtccttcctt	tggatccctg	cagtgtgtgt	gctcagcggc	ttgcacgtag	acttgcctag	51480
aagaaatgca	gagccagcct	gtgtgcacca	ctttcagagt	tgaactcttt	aagcccttgt	51540
gagtgggctt	caccagctac	tgacagaggca	ttttgcattt	gtctgtgtca	agaagttcac	51600
ctttctcaagc	cagtgaataa	cagacttaat	tcgtcatgac	tgaacgaatt	tgtttatttc	51660
ccattaggtt	tagtggagct	acacattaat	atgtatcgcc	ttagagcaag	agctgtgttc	51720
caggaaccag	atcacgattt	ttagccatgg	aacaatatat	cccatgggag	aagacctttc	51780
agtgatgaact	gttctatttt	tgtgttataa	tttaaaactc	gatttctcta	tagtctttta	51840
agttgacatt	tctgcttact	gctactggat	ttttgctgca	gaaatatatc	agtggtccac	51900
attaaacata	ccagtgggat	catgataagc	aaaatgaaag	aaataatgat	taagggaaaa	51960
ttaagtgaact	gtgttacact	gcttctccca	tgccagagaa	taaactcttt	caagcatcat	52020
ctttgaagag	tcgtgtgggt	tgaattgggt	tgtgtacatt	agaatgtatg	cacacatcca	52080
tggacactca	ggatatagtt	ggcctaataa	tcggggcatg	ggtaaaactt	atgaaaattt	52140
cctcatgctg	aaattgtaatt	ttctcttacc	tgtaaagtaa	aatttagatc	aattccatgt	52200
ctttgttaag	tacagggatt	taatatattt	tgaatataat	gggtatgttc	taaatttgaa	52260
ctttgagagg	caatactggt	ggaattatgt	ggattctaac	tcattttaac	aaggtagcct	52320
gacctgcata	agatcacttg	aatgttaggt	ttcatagaac	tataactaat	ttctcacaaa	52380
aggtctataa	aatacagtcg	ttgaaaaaaa	ttttgtatca	aaatgtttgg	aaaattagaa	52440
gcttctcctt	aacctgtatt	gatactgact	tgaattattt	tctaaaaatta	agagccgtat	52500
acctacctgt	aagtcctttt	acatatcatt	ttaaactttg	tttgtattat	tactgattta	52560
cagcttagtt	attaattttt	ctttataaga	atgccgtcga	tgtgcatgct	tttatgtttt	52620
tcagaaaagg	gtgtgttttg	atgaaagtaa	aaaaaaaaat	aaaatctttc	actgtctcta	52680
atggctgtgc	tgtttaacat	tttttgacct	taaaattcac	caacagcttc	ccagtacata	52740
aaataggctt	aatgactggc	cctgcattct	tcacaatat	tttccctaag	ctttgagcaa	52800
agtttttaaaa	aaatacacta	aaataatcaa	aactgttaag	cagtatatata	gtttgggtat	52860
ataaatccat	ctgcaattta	taagatgcat	ggccgatggt	aatttgcttg	gcaattctgt	52920
aatcattaa	tgatctcagt	gaaacatgtc	aaatgcctta	aattaactaa	gttgggtgaat	52980
aaaagtgcg	atctggctaa	ctcttacacc	atacatactg	atagtttttc	atatgtttca	53040
tttccatgtg	atttttaaaa	tttagagtgg	caacaatttt	gcttaatatg	ggttacataa	53100
gctttatttt	ttcctttgtt	cataattata	ttctttgaat	aggtctgtgt	caatcaagtg	53160
atctaactag	actgatcata	gatagaagga	aataaggcca	agttcaagac	cagcctgggc	53220
aacataatga	gaacctgtct	acaaaaaaat	taaaaaaaat	tagccaggca	tgggtggcgta	53280
cactgagtag	tttgtcccag	ctactcggga	gggtgaggtg	ggaggatcgc	ttcagcccag	53340
gaggttgaga	ttgcagtgag	ccatggacat	accactgcac	tacagcctag	gtaacagcac	53400

gagaccccaa	ctcttagaaa	atgaaaagga	aatatagaaa	tataaaattt	gcttattata	53460
gacacacagt	aactcccaga	tatgtaccac	aaaaaatgtg	aaaagagaga	gaaatgtcta	53520
ccaaagcagt	atlttgtgtg	tataattgca	agcgcatagt	aaaataattt	taaccttaat	53580
ttgtttttag	tagtgttttag	attgaagatt	gagtgaaata	ttttcttggc	agatattccg	53640
tatctgggtg	aaagctacaa	tgcaatgtcg	ttgtagtttt	gcatggcttg	ctttataaac	53700
aagatttttt	cttctctctt	ttgggccagt	tttcattacg	agtaactcac	actttttgat	53760
taaagaactt	gaaattacgt	tatcacttag	tataattgac	attatataga	gactatgtaa	53820
catgcaatca	ttagaatcaa	aattagtagt	ttgggtcaaaa	tatttacaac	attcacatca	53880
ttgtcaataa	ttcatgtaat	taactgaatt	taaaaccttc	aactattatg	aagtgtctcg	53940
ctgtacaatc	gctaatttac	tcagtttaga	gtagctacaa	ctcttcgata	ctatcatcaa	54000
tatttgacat	cttttccaat	ttgtgtatga	aaagtaaatc	tattcctgta	gcaactgggg	54060
agtcataat	gaggtcaaag	acatatacct	tggtattata	atatgtatac	tataataata	54120
gctggttatc	ctgagcagg	gaaaagggtta	tttttaggaa	aaccttcca	aatggaagc	54180
tgaagtactt	ctaataact	gaggggaagta	taatatgtgg	aacaaactct	caacaaaatg	54240
tttattgatg	ttgatgaaac	agatcagttt	ttccatccgg	attattattg	gttcatgatt	54300
ttatatgtga	atatgtaaga	tatgttctgc	aattttataa	atgttcatgt	ctttttttaa	54360
aaaagggtgct	attgaaattc	tggtgtctcca	gcaggcaaga	atacttgata	aactcttttt	54420
gtctctttat	ggtattttca	gaataaagtc	tgacttgggt	ttttgagatt	attggtgcct	54480
cattaattca	gcaataaagg	aaaatatgca	tctcaaaaat	tggtgataaa	aagttatttc	54540
ttgtatatgt	gataaagttt	acatgttgtg	tatatatgtt	gtattgccaa	atcgggctat	54600
taaatactac	gtcatatttt	aaaggttcag	ttttagtgta	tagtaaacaa	gcagtgcact	54660
aagcctcttg	cgggcatcat	ctcatctcac	tgatcatcaca	aaccccatgc	cacagcgtag	54720
cttgaccact	aaaagtaatg	catctgcgaag	catactgcc	ggttttggat	agtttgtacc	54780
aacagttacc	ttatcaaggt	aaatcccaga	ctctaaaaga	gttgggtgctg	tgctactaca	54840
tgcataactt	taaataaatt	tcctgccggg	cgcgggtggct	cacgcctgta	atcccagcag	54900
tttgggaggc	cgaggcaagt	ggatcacttg	agggtcaggag	tttgagacca	gcctggccaa	54960
cgtggtgaaa	ccctgtctct	actaaaaata	caaaaattag	ccaggcgtgt	ggtggcaggc	55020
acctgtaatc	ccagctactt	gggaggatga	ggcaggagaa	tcatttgaat	cctgcaggcg	55080
gaggttgtag	tgagccaaga	tgccgtcatt	gcactccagc	ctggggcaca	agagcgagac	55140
tccgtattaa	aaaaaaaaaa	aaaaaaaaaa	aaaaattcct	ctcctgtttg	agctttccct	55200
tacctgtaaa	gagggggagaa	tatgtattta	cttcaaagag	ttcagggaaa	tgactctcac	55260
tagtttgaga	ttttaggtat	aaaaatacat	tcttatataa	ttttaacacc	aatgtgagag	55320
attattattc	ttgttaaacc	aattcagttt	tatttgcgtg	ctaaaaatgtg	tgaataagta	55380
attgtccatt	attttctgaa	gtgttttggg	actcaacaca	tgattgtgag	gaggatttgt	55440
tgctaaacat	ctttctgggt	attcaagctc	gtgtatactg	tgctctgttg	agacatgcag	55500
agttactttc	tgctctgggtc	acaggtcagt	tctgtagagt	tttcggacaa	ttaaccagtt	55560
ttcatttgcc	catgaccacc	tttattcttt	ttcctcaact	gcacccatct	tttataaggt	55620
ctttcagttt	attgcagaga	agatgggtgga	gaaaagccgg	aattcccacc	caccgctgcc	55680
atccccatgt	tttatcattg	gctagagtgg	aaaatagcag	taactactgt	gagagatcat	55740
ttgtttatat	aatggaaaca	aagatgagga	agaacactgg	cttagatcac	agaactgatg	55800
tattttagatt	cttttttttt	ttttttttta	gacggagtgt	tgctctgttg	cccagactgg	55860
agtacagtgg	ctcaatctcg	gctcactgca	acctccattt	ccctgggttca	agcaattatc	55920
ctgcctcagc	ctcccaagta	tttgggatta	caggcgtgtt	ccaccacacc	tggtctaattt	55980
tttgattttt	tagtagagac	gggggtttcgc	catgttggcc	aggctggtct	cgaatcctg	56040
acctcagatg	atccaccgcg	cttgccctcc	caaagtgtg	ggattacagg	cgcgagccac	56100
cgcgcctggc	ccaatgtatt	tggattctta	aagaacactt	tcaaattaaa	tatcagttga	56160
agagaactag	aactaaagaa	tttctgtgtc	aaactgttta	gcaaatgtaa	gtagaagctg	56220
ggagatgtgt	cctggaatga	atgaatacat	cagtaaaata	ccatacgtat	gttatgatgt	56280
tattgtttcc	ttgccttggg	tgatttgggt	ttactgtgaa	ataattttca	atatagaatt	56340
gtgatcgttg	gaatttgggtc	atctactaga	aaatgagaaa	gaagttaata	gctatcttcc	56400
ttaaaagattt	ctgaggttgg	gattaaggtg	gtgttcccaa	ggtgttctaa	aacggcagcg	56460
agagctgtgc	actcacttca	caaatttgaa	ttcctgctct	gtgttaggcg	ctgtgctagg	56520
<210>	180					
<211>	2000					
<212>	DNA					
<213>	Homo sapiens					
<400>	180					
gtggatctgt	gactgttcgc	aggaagagag	gagcgggagc	aggacagaca	ataactgata	60
gtcaggagct	gggttttgag	ataaagaggg	aacaagagaa	agtttaagttc	tggtttttca	120
tggaacacat	tgacacaaa	tttacaactt	cgtgactaac	agtaactctg	ggtgattcac	180
aacaaattta	carataaaca	catatttact	gactttatatac	acagcaatcc	taacgtgaac	240

acagaacctg	ctttatcttt	tgcacactg	ttctagtgtg	gagatgtctg	gtctcagtta	300
aagaaagcat	aaggagcatt	agttgtgcac	actgtccaca	cccgtgacct	ttttccacca	360
gtactaaacc	tagtgcttct	tacagtacag	ggcaatgaca	gccacagaaa	gagagaagct	420
ccttttactg	tgtaatgctt	cctgctggcc	ttcaaatact	tgttacttga	gagatctcca	480
ttcacctggc	tttgtcccca	aaggctcatca	tctaccaatg	atgttggttat	ttgatgttaa	540
tcattgtataa	agaaagtagc	taccatcctg	gccttgatta	gaacttccca	ctgaaatacc	600
gtcctgccta	aaggtagcac	aggtttccat	tatggtggtg	gtggggaggg	ggcggaata	660
tatatatata	tatatatata	tatatatata	gtaaagcatt	cggcattctt	ttaaagtaca	720
actatccttg	aaaagggtta	catattaaac	cattttttacc	acagccaaag	gggaggagaa	780
agatccaaaa	gtcctgtgga	tctgctttaa	catcaataaa	acagttatcc	acccttcgta	840
gcttttagtg	aaggctacaa	aagtatgctt	tttatggatt	acacatgtgc	acgcaactac	900
tttaattact	acagaaaaaa	acgaggctcc	ttattaaaaa	aaaatcagaa	acaagtccaa	960
cagactctga	ggaaatgaag	caagagtga	ttctgaaaaa	gtctaataaa	cagtatggaa	1020
atatacctgt	gggattgttc	ttcagctatg	cataaacatg	taattatcat	cattactgtg	1080
atggggaaaa	acacgggacc	taattctgaa	acaccctggg	agcgagagac	ggcgaggagg	1140
ggctgtctgc	cactcagagc	ggagggtgag	gaggcggtct	ccccttgcaa	aggactggca	1200
gtgagcagat	ggggacactc	gagctgcccc	gcgacctggg	ccgagctgcc	tacaacctgg	1260
gccacgtgct	ctgcaagaat	tagacctccg	ataacgttaa	caccaccttt	ctcactgtcc	1320
taattgtgtg	catcccggcg	cccaggggct	tgtgagcagc	aggtgcgcgt	tccaggcagc	1380
tccagcgacc	cttaaacctg	accgcgcgca	cgcccgcccc	gagggagcag	aacaagagcg	1440
acccggaccc	tcctccggcc	agcaccaccc	ttcaccaggt	tcctcagtc	gccaccacct	1500
cccttccccc	gtccgcagcc	ggcccagctg	gggagcatgc	gcagtggccg	gagccgggtt	1560
gcccgcgcca	cagcaggtag	ctgtactgca	actgtcggcc	caaaccaacc	aatacaagaga	1620
cggtgttattg	ccgcccaggt	ggaactatgg	caacggggca	ccaatcagaa	ggcgctgtgt	1680
tgcgcgggag	ccccttgccc	cggcaggggg	atgtggcgat	gggtgagggg	catgggtgtg	1740
gagcatccct	gagccatcga	tccggggagg	ccgcgggttc	ccttgctttg	ccgcccggag	1800
cggcgacagc	agccccgcac	tcgcctaccc	ggccccgggc	ggcggcgcgg	cccatgcggc	1860
tgggggcgga	ggctgggagc	gggtggcggg	cgcgccggcc	cgggcccggg	cggtgattgg	1920
cgcctgctg	gcgcgactg	aggccccgga	ggcggcgggg	gagcgcaggg	ggagctcgct	1980
gccgcgagc	tgagaagatg					2000
<210>	181					
<211>	1901					
<212>	DNA					
<213>	Homo sapiens					
<400>	181					
taaaggttca	gtttgtagt	atagtaaaca	agcagtgcac	taagcctctt	gcgggcatca	60
tctcatctca	ctgtcatcac	aaaccccatg	ccacagcgta	gcttgaccac	taaaagtaat	120
gcattctgca	gcatactgcc	aggttttggg	tagtttgtac	caacagttac	cttatcaagg	180
taaatcccag	actctaaaag	agttgggtgt	gtgtcactac	atgcataact	ttaaataaat	240
ttctgcccgg	gcgcggtggc	tcacgcctgt	aatcccagca	gtttgggagg	ccgaggcaag	300
tggatcactt	gaggtcagga	gtttgagacc	agcctggcca	acgtggtgaa	accctgtctc	360
tactaaaaat	acaaaaatta	gccaggcggt	tggtggcagg	cacctgtaat	cccagctact	420
tgggaggatg	aggcaggaga	atcatttgaa	tcctgcaggc	ggaggttgca	gtgagccaag	480
atggcgctcat	tgactccag	cctgggcgac	aagagcgaga	ctccgtatta	aaaaaaaaaa	540
aaaaaaaaaa	aaaaaattcc	tctcctgttt	gagctttccc	ttacctgtaa	agagggggaga	600
atatgtatct	acttcaagaa	gttcagggaa	atgactctca	ctagtttgag	attctaggta	660
taaaaataca	ttcttatata	attttaacac	caatgtgaga	gattattatt	cttgctaaac	720
caattcagtt	ttatttgctg	tctaaaatgt	gtgaataagt	aattgtccat	tattttctga	780
agtgttttgg	aactcaacac	atgattgtga	ggaggatttg	ttgctaaaca	tctttctggt	840
tattcaagct	cgtgtatact	gtgctctggt	gagacatgca	gagttacttt	ctgtctgggt	900
cacagggtcag	ttcttgatag	ttttcggaca	attaaccagt	tttcatttgc	ccatgaccac	960
ctttattctt	tttctcaac	tgcacccatc	ttttataagg	tctttcagtt	tattgcagag	1020
aagatggtgg	agaaaagccg	gaattcccat	ccaccgctgc	catcccatat	ttttatcatt	1080
ggctagagtg	gaaaatagca	gtaactactg	tgagagatca	ttgtttata	taatggaaac	1140
aaagatgagg	aaagaacctg	gcttagatca	gagaactgat	gtatttagat	tctttttttt	1200
ttttttttta	agacggagtg	ttgctctggt	gcccagactg	gagtacagtg	gctcaatctc	1260
ggctcactgc	aacctccatt	tccctggttc	aagcaattat	cctgcctcag	cctcccaagt	1320
atttgggatt	acaggcgtgt	tccaccacac	ctggctaatt	ttttgtattt	ttagtagaga	1380
cgggggttctg	ccatgttggc	caggctggtc	tcgaaatcct	gacctcagat	gatccaccgc	1440
ccttggcctc	ccaaagtgtc	gggattacag	gcgcgagcca	ccgcgcctgg	cccaatgtat	1500
ttggattctt	aaagaacact	ttcaaattaa	atatcagttg	aagagaacta	gaactaaaga	1560

atttctgtgt	caaactgttt	agcaaatgta	agtagaagct	gggagatgtg	tcttggaatg	1620
aatgaatata	tcagtaaaat	accatagcta	tggtatgatg	ttattgtttc	cttgcccttg	1680
ttgatttggt	tttactgtga	aataattttc	aatatagaat	tgtagatcgt	ggaatttggt	1740
catctactag	aaaatgagaa	agaagttaat	agctatcttc	cttaaagatt	tctgaggttg	1800
ggattaaggt	agtgttccca	aggtgttcta	aaacggcagc	gagagctgtg	cactcacttc	1860
acaaatttga	attcctgttc	tgtgttaggc	gctgtgctag	g		1901

<210> 182

<211> 4550

<212> DNA

<213> Mus musculus

<220>

<221> exon

<222> 2259..2488

<223> exon1

<400> 182

tacagctatc	tgtgtgtatg	ggtgcatgcc	atggaccacc	tgtagaggtc	acaggacaat	60
ctccaattca	ctttctcctt	ccaccacatg	ggctctgtaa	tcactaaggt	ttgtcaatcc	120
tgagtacaga	tgttcagaac	catcttactg	tctcctctct	tctgataaaa	catgaggttg	180
cccagagac	gttttagaca	gggttataat	ctgataaagg	aaaagccaca	tgctctttcc	240
ttacaaatgt	aatttctaca	gacattccta	gaaaattgaa	actttatggg	tgaggaaagga	300
gagggggccc	tcaggtacct	tgtttttctg	ttgacaaaag	ttgactctta	acattgtcaa	360
gtaaaatgctc	ccacaaatgg	atcatctgac	tatttgcaga	atgtcatagg	ccaacagaga	420
gagaacccct	gaatttccag	agaccttcag	gttggtctcag	tcccttcttt	tttgaatgtg	480
acctcaattc	ctgtcttctc	gaactcttgt	ttgccaatct	gaatctacag	tctatctgtc	540
aaacaattcc	tttgtctgga	ctggctctgct	gaactgacag	tgaattgtct	tgacagttcc	600
tttgctgcc	cttttacctc	tgcatcttca	ttaaactgga	cagtttgtca	tatctgtgac	660
ccaccaacag	ctgcttttcc	cctaaagctg	ggtttgtggt	tcatgttatc	gtgacagaca	720
ctcttatagc	cctgtcagtt	ctccagcact	ggcttcccaa	ggcttttaaa	actcctttct	780
tctttctaac	tctttgtagt	cactgttaacc	tatatatgca	tatgtaaaca	gagatatact	840
tacagagtga	tgtatgtgtg	atctgagagt	taatattagt	aattaagact	gcaataaaaag	900
aaactgtgtt	tcccttagca	agggtacag	agtaaaagtg	gcctctcttg	tgccagcgaa	960
gccactgtac	ttagtgaat	ttattgtcat	tcaatacatt	ctgatattct	gtaaactcct	1020
aagcacgtcc	atctgacata	gtgtgctaat	gacaggagtc	acctgtatgc	cttatgaagc	1080
gcatctcaga	ggtgatggga	aagaaacatg	gggcaaaaga	tgaagggaaa	tccaaggcaa	1140
ggaagcagag	acacaggcgt	cagtgggtgtg	gaaagggaga	aaactagggg	cagaataagt	1200
gaccttaggg	tcacttagag	aaaccaacac	acacacacac	acccacatat	ttaaaacgta	1260
ctttatacac	atctgagcgt	gcgcactgac	ctgtttcctt	ctataacctc	ttgtatagaa	1320
ttatctggtc	tccactagtt	agggcagtga	aaggacctgg	gccccctggat	aagtttttgc	1380
tgttacttaa	ctattctagt	tttctggagg	gaagagaact	tatggatcct	acatgtatag	1440
ggaaatactt	tcctacacat	tgaagaagaag	aaatgtagga	tattagggaaa	acgcacagta	1500
gaaacaagtt	aaagagcaag	aggttattaa	agggcaaaag	ttaaggcttt	gaaagattta	1560
atacaaggag	gtgacagtcc	cgtgaaaggt	gaaccaaggg	tacaggagac	ggacccagcc	1620
tcattctgca	acagccaaga	ggaggggaagg	tggtgttctt	atgcacgtgg	gggcacgggt	1680
ggccctccgg	cacgcgaaga	cgctgcagtt	gtccataacc	tgccgcatcg	agctcctcct	1740
gtgctccacg	acttagtcgg	ctcacgcgtg	tcttgacagga	agcatcctcg	tgctctccacg	1800
cagctctcgc	acgccagcac	aggccaaaac	ccaccacctc	acttcttccc	gggctcatcc	1860
ccagccagca	ttcgagtcg	agcatgcgtc	gtgacgaggc	caagggaccg	agccaatcag	1920
aacacgtatt	acgcccataa	gtcggccaat	caggaggcgc	cttattacc	gggagccttg	1980
cttcaccccg	cctccccgct	gacaagcacg	ggtcgcgcgg	agcaaaagca	gcaccccgag	2040
gcgagtgcgc	ccggcaagcc	gaggcgtgcc	ctttccaagg	cgccgagcag	aggccgtcac	2100
tgtccccgcc	gggtccccgg	cccccgcggc	ccatgctggg	ggcggagcca	gggcggaggg	2160
ggcgggcgcg	gccggccccg	cgcagtgatt	ggcgggcggc	cgccggtggc	tgaggtcctg	2220
gtggccgcgc	gggcaacgca	ggcgagtcg	cggctggcga	gccgagagga	tgctgctgtc	2280
cctggtgctc	cacacgtact	ctatgcgcta	cctgctcccc	agcgtcctgt	tgctgggctc	2340
ggcgcccacc	tacctgctgg	cctggacgct	gtggcggttg	ctctccgcgc	tgatgcccg	2400
ccgcctgtac	cagcgcggtg	acgaccggct	ttactgcgtc	taccagaaca	tggtgctctt	2460
cttcttcgag	aactacaccg	gggtccaggt	gagggcgggc	cgcgaggggc	tgctgtcgag	2520
ccctccccgc	ggccggggcg	gcgcttgcaa	ccggggcgaa	cactcgagc	ccggcgagca	2580
cgtgcccag	ctcacggcct	cccgcgcggg	ggggaaagtt	ctgggttctca	cttcgggggt	2640
ccttctggaa	cgctcctgctg	aggctgagtg	tggtccccgg	tccgccccac	ccccgccccg	2700
ggccggctgt	tactgcccac	ctcagtcgct	gccaaagtag	ggcactgagt	ccgaggtggt	2760

gatgctggga ctggcttcat ttgcacttcc gaggtctttt agattagcaa gacctctagg 2820
 cgctgaccaa agtgacagct gtgaaggacg actcctgcct tgggttcctc ccgggtgaaa 2880
 gcgagggcct agggaggaaa tgaatacatt ggttacaata ggagcctcac tgcgataca 2940
 gttctcttca gcttggactg ggcttcaatg tgggttgatc tcttgtcaga ttgctttctt 3000
 cctgctactg tttctttctt tctttccanc cctccctccc cccccccc cgccccgtgg 3060
 agattgaact ctgaaaacaa taaagagtag aaagctctcc taatgtgaat tcgttatatg 3120
 acatcccata aaaacctaca gttgtacttc ctttttgggt ttcagtttca aagaagagct 3180
 ctggttgggt tctcccagat gtatctatga ctttcccccc catttctcag ttcttttcat 3240
 tctgtggttag ggggttactt tggcgactgg atcccttact gagttttgcg ccagttggag 3300
 attatgtctg aggtagggaa ttaagacctc tctgaatcac tatcttttta aatgttttcc 3360
 taggggaatag gaaaatcact gttgcacatc aaggtttctg aaaaattgac ttttagaata 3420
 ggatttcatt cagaattttt aggaaccccc acactgatgg tttcaaacct ccctcttact 3480
 ttactaagtt tgtcaagtga atgtatgggtc taatcgtgga taagtattta atttcactag 3540
 cagaagggac aagacagcgg ggagcacaac ttaaggttgc tgaccttgca catgacaagt 3600
 acccctcaga cgctcaggga cctctactca agtgccacct atattcttgc tgcagagacg 3660
 ttaggatgag tcagaatgaa gcaaagtttag tgagtttatt gattggggaga gaggacacgc 3720
 acttgagggg agtcaagtgc aaaccttatt accccccacc caggctacag cagctgtttt 3780
 ctaagtgatt ttagggtctt taagttaacg ccttaaaact aagattaagg agaagagaag 3840
 gaaaaaaatg agttcttcta ttctttccaa taatgagctc taaaaaaaaa agaagcaaac 3900
 caggatctca cactgtagtc ttggtgggca ggaactctat gtagacctca caggcctcaa 3960
 gttcacagag atctgcctgc ctctgtctcc agagtgttag gactaaaggc atgtaccgcc 4020
 atgtctggat taaactcttt tagttatatg aaatttaaaa cggatttcag gcggtagtga 4080
 acagttttaca tatgaggggg aaatgtgggt aggcagtaat atggatcaaa ataaaatcaa 4140
 agtaattagc tgatcactgg tcacaagagt ttgagatgtg agcttgtctt ctgccttagg 4200
 tcaccagcta tagggataat cttttgtttg tttttgttgg tttttgttgg tttgtttttt 4260
 tgtttttttg agacagggtt tctctgtgta gtccctggctg tcctggaact cactctgtag 4320
 gccaggctgg cctcgaactc agaactccac ctgcctctgc ctccaagtg ctgggatgaa 4380
 aggcgtgcgc caccacttgc ctataatctt acttgtaatg gttttagaat atgtgcacag 4440
 tggagagcag tgttcaagca gctgtatcca accaatttca cttaaagagg gagagggtga 4500
 ggggtgagggc ctctcttttc tattcaaaag cagatttgtt ggacattgca 4550
 <210> 183
 <211> 37950
 <212> DNA
 <213> Mus musculus
 <220>
 <221> exon
 <222> 5259..5828
 <223> exon2
 <221> exon
 <222> 12675..12791
 <223> exon3
 <221> exon
 <222> 14621..14710
 <223> exon4
 <221> exon
 <222> 19822..19912
 <223> exon5
 <221> exon
 <222> 21789..21950
 <223> exon6
 <221> exon
 <222> 23387..23510
 <223> exon7
 <221> exon
 <222> 25520..26016
 <223> exon8
 <400> 183
 tggagtgaga ggcctgggta taattccttt ttctttgtca cactgtagca gttctgcttc 60
 tcagcctcag ttgagactgg aatacatttg tcatgctgtt ctgaagactt taatggttga 120
 tctttactgt caccttgact ggatttagaa tcgccttgga gatgtgctta tggatgtct 180
 gggaggatgt ttccagaaaag tgtttactga aggtgtcctc atctgatagg gtggggccct 240

ggactcaata	aaaaccggca	tttatctgtt	tcctgtccag	ggacacagtg	tggccagcta	300
cttcacactc	ttgctgcctt	cctacatcga	aagactgtat	cttttcttaa	aaggcgaggc	360
aaaataaatac	cccaccccca	ggattttttt	tggctgtcct	ggaactcagg	tccacctgca	420
tctgcctcct	gagtgtctgg	tctacaggag	taccacaccag	gcctggccta	aagaaacctt	480
tcttaaatgt	ctcctgtcag	acactttgac	acaacaataa	gaaaaataat	tcatacaaa	540
accaagttag	ataaatatgtt	atttttttag	ttcatagagt	agtaggtaat	ccatgggtgag	600
ggaaaaaaa	aaaaanaacc	cttttgaat	aaggacagta	ttgagaaaca	tgatgtctga	660
ttgtatgagt	tttgtgatat	gtaagtttct	actcaattca	catggtaatt	gtgcattctg	720
atcataatc	aaataattgt	gtttactact	ttaaccttct	tacaaagtat	agttttacta	780
attagtaatt	tagtaaat	attatttatt	taatgaaaca	tttcaaattg	gactcagaaa	840
aagaccacag	tttttgtaca	tattatagta	gaagtccctt	agtaggtgga	aatcctgtgt	900
ttctttacaa	ggatatgtct	agaacacgtt	aaacaaacag	gaggaggtgt	ggctgcaccc	960
gttaggctaa	ccagtcacaa	tgctttttaa	agccatacgt	gttgtgtgtg	agcatttttt	1020
taaatatata	gaaaatcccc	aaaatagcta	gtataataag	cacacatgcc	agtaagcctc	1080
ttattactgt	aaaataactgt	gtaatacttt	gtgcgttctt	ttatgtgact	gcagtaggtc	1140
tgtttaccatc	agcattttacc	catacaaggt	ggctactgtc	attcaggcct	tgaaagttt	1200
tcagcttctt	tgaaacttgt	ggggttactt	ctccagttgt	atgtgtgtgc	catgggtggc	1260
tagacatggc	acatgactgg	cgcattggacc	ttaaagacaga	ttaaagaatt	catttcaaat	1320
gatttcaagc	acagagttta	aattgtacat	gcactcttag	tcaatgtcct	tgtagctgca	1380
ttttgtgtga	attggagggg	catggactag	ttagctgttt	cttttactgt	atgacagttt	1440
tgaatgaaaa	gcattgttaga	tacaaaataa	tttaactgtt	gacccccccc	cacacacaca	1500
cacacactct	ttttccctgt	atgccttcac	tcctcgaatt	ggtttttctc	anatagctgt	1560
tccagtgta	ccaggcatat	atgaagtagt	ttatgtagtt	tcattttatg	cagtgcacac	1620
ctgtgaggag	caattagtag	attatacttt	atagttaatg	agatagatat	agagaaagct	1680
gatacatcac	atctatttct	tggtgaagat	ggagggtctg	gattcaaatg	taagtcttag	1740
tcogtgcctt	atactttgca	gcctgtgtgc	ttagtgtatg	tcattgttatg	aagggcaacc	1800
tccttttccag	actgtgtctg	caagaccctg	gaattaaatg	accaaaaggc	tactaactgt	1860
agagaaactg	cccattttat	tgatgcta	atttttaccat	ggtagggaga	aacttggaag	1920
aaatgagaag	ccctactcag	ggggattttt	caggtgagat	tatgtagtga	ctagtgtaaa	1980
agaagctact	ataaggagata	ccaagtatg	gaaataagtg	ctgcacacct	caggtgtgtc	2040
acacagactc	agactgacag	ctcagggtctt	cctgtcgaag	aaggggacgt	ctttgaaagt	2100
gagaaattca	tgtcttcttt	atagaaagt	ctactccagc	tctaggccga	agactggggg	2160
cagagagctt	ttcttgtgcc	tgcgatttct	taactgtttt	tattcaaaat	cattcttatg	2220
ccaaaagggc	atatttgggg	ttgagttctt	tcagcgatat	catttatatt	tgagaccaat	2280
gagacgtttt	tctcactatg	tattgtattt	caactttcat	gctaaacctt	ctggacttta	2340
ctgagataaa	tcaagaacat	acctttaaac	ttttagtttc	ttctttgcca	ctgtgaccaa	2400
aacacatgac	tggggagaaat	ggtttatttt	gcctcccagt	cttgggtgatt	ttagtccctc	2460
acagcagaga	agcctggtag	agcagctcac	tgagtggcaa	caggactgtg	ctcatgtgat	2520
catggaccag	gaagtgtaga	acaaagccag	aactaggggc	cctagtaacc	tacttctgcc	2580
agctaggccc	cacttccctga	agggtccacc	tcctccctcc	ccctcccaaa	aaagaaaata	2640
aaaatagtag	caccaactgg	acagcaaaca	ccccaacat	aagccagtga	ggaggggaag	2700
aggagggggg	gaggaaggag	ggagagaggg	gggaaggaga	aggagagaga	gaaggagaga	2760
gagagggagg	gagggagaga	gagagagaga	gagagagaga	acacaccanc	tctcaagtat	2820
nnnnctnna	nnncttgga	caaattaaaa	actatgtaac	cagaaaatta	ttttaagagt	2880
atttgatttg	tctgtattta	attaattaac	tgaaaaaaa	aggaaaatta	attcctttct	2940
ctaaagactt	taaatgaacc	atttttttag	tgtagtgtgtg	tgtagtgtct	ggtaaaaaga	3000
caggtctcag	gagttgat	ttgacctgt	gaacctagg	gatcgaactt	ttagctctc	3060
agctttggca	gcaggaatct	ttatacactg	catcatctca	ctggccttag	ataagttttg	3120
aaaaaaagt	gaagaatcta	aagttacttg	gataattata	taaaatataa	gtctgaggtt	3180
gggctcaaga	cggaaagaca	tcagtaagga	gccaagagaa	cagccccaga	gagaatctga	3240
gaatcaagg	tggtggcaaa	catacagtga	tctacacttt	ttatctgtaa	atagtttgta	3300
aattcttaac	cctttaaaaa	aattccctaa	cccctactct	gcagatcaga	ctgccctgga	3360
actcattgta	gaagcctaag	gtggcctcgc	attcacacaa	tccttctgac	acagctcccc	3420
aagtgcagg	attacaagga	taagctactg	tgctgtcctt	cttagctttt	taaattttaa	3480
aagcactaca	acpttgtaac	tagttttata	cttgtaata	aaataacagt	aactgggact	3540
ggggagatgg	ctcaggctgc	ttttccagag	ggacccaggt	tcgattccca	ggagccacac	3600
aatggcttac	aatcatctgt	agctatagtt	cccagggatc	tgatgccttt	ttctgattcc	3660
tgacggcccc	aggcaagcat	ggggtgtaca	aacacacatg	caggcaaaac	agatgtcttt	3720
ctgcattgct	cttcaccttt	tattatgagg	caaggtctct	cacttgaacc	cagatctcac	3780
tgattggcta	gtgtaactaa	ctggctgtct	caaggaaatc	ctgtctacac	ttcactaaag	3840
ctcttgctgt	gragcccagg	ctagcctcaa	attctaatc	ctcctgcttt	agccacttaa	3900

gtgtccggga	tttcaggcat	gcaccactac	atctgggtga	aaggatatctt	tggatgggttt	3960
gggtattatt	agcttggagg	ctagcgatgt	ctccaaagaa	ccttacataa	tttacataatg	4020
catacatata	tatatata	aacatacata	catacacaca	tgcatacatt	taactggatc	4080
tgtagtctct	gtaaagtgt	cccagtggtc	cttaaatagt	tctactttat	ttctaagagt	4140
gatcatgac	atgagctggc	atcccaatat	taactctgca	cagatcaact	aaccattggt	4200
tacttatttg	tttgatttat	gtcgtttgta	aacttgatta	gaagtaatta	gacacgtgtg	4260
aacacactgt	gcagttgtct	tagagcagta	gcagcagcct	tagtgtgagt	ccttgacatc	4320
taactttata	ttatgtagct	acattgcttt	tacacagggc	ttattaatgc	cctttcagggt	4380
tttgttatca	ttatgggtatt	tgtgagtatt	tcaccaccag	gaaggcaaac	ctttctggca	4440
tagtagactc	caaaagcgtc	actaaattat	gcataattct	tagtgacaaa	gtataagaac	4500
aactgctaag	taaaagtgat	ccaaaggaag	ggaacagatg	agcaacagca	gtcatttgct	4560
cccttcaagt	ccctttctaa	tcactggata	cttaacactt	acaaaacaga	acagtacagt	4620
gaaaacctgg	cctcgccct	tgggttaaaa	acagagtgt	gccaggaggg	tacagtggca	4680
catgatgtag	tgaccttgac	tgatgttcc	agctcacagg	gacctctggc	ctccctgttg	4740
tactccctg	gctccttagc	ctgcatTTTT	gttcagcaca	caagtggact	aacacaagac	4800
tcagcacac	agcttctatt	tggctcttatg	ctcagaaac	ctttcctctc	atttgctaaa	4860
atataatcaa	cgacacagct	agctctccat	tttacaatc	cttaaaatat	agagtatgtt	4920
cataatattt	gcgttatttc	ccattatgtt	attatgtata	tgcccatctc	ttccgctgga	4980
ctgtagagcc	ttatatatta	ttagtatca	ataacttctt	gtggattttt	cttacaggga	5040
tatcaatcat	tctgtcttac	tttatgatca	tttaacctgtc	ctacatttaa	gtactttata	5100
agtatgagct	atttatagt	tgtggggctc	ccaagagaac	tctgatgttt	gttagtcatg	5160
gataagctta	gttacattgt	gtccttcttc	gtttcctttc	acattctttt	ttttttttta	5220
attgtcaaag	tcatgattct	ttttgttttc	tttttttagat	attgctatat	ggagatttgc	5280
caaaaaataa	agaaaatgta	atataatctag	cgaatcatca	aagcacaggt	ttgtatttca	5340
tttgatgaaa	tttgggtttt	tctagaaatg	gtaaatgagc	attaatatgt	acacacacat	5400
acacacaaac	acacatatgt	acacacacat	atgtttttaa	gacaggattt	catgtgaccc	5460
agaatggcct	catactctct	gagtagctga	gaatgatttt	aagcttggtga	cacacctgcc	5520
ttcatctcca	aggtacagga	attgacaggtg	ctttctttgt	atgcttagct	tatgtgggtc	5580
ggggcatcaa	acccaaggct	tcattgcaaac	taggcaagca	ctgtcccatc	tgagctacat	5640
ccccagccca	tcaaaatgat	attttaagggt	tatttattta	atgagtttta	tcacgtgtgt	5700
gtgtatctgt	ctgtgggttt	gagcttggtga	gtacacatgc	ctgtggaggc	cagaagaaga	5760
acaccagatc	ttccctggag	cttgagttgc	aggtagcgat	tagttatcct	ggatggattt	5820
ggggaactaa	actgggggtg	tttgaaagag	aaatatgtac	tcttaactgc	tgggctttgt	5880
ctccagcctt	taaaatatta	atcttatata	tttaagttaa	ctaagctagc	ttttgttttt	5940
aacataaatt	tgctgtggat	tttgaatctg	gcttgcaatt	ttattttact	tttttggtgg	6000
ggaggggtaag	gttagtaagt	tgaagggtat	ggtttttgtc	caggcttctc	ttcttcctta	6060
tttctcaaaa	taacccttta	gtttatttgg	ttttctgtct	ctacgttata	ttttctagaa	6120
tatatatata	tatatacaca	cacacacaca	cacacatata	tatacatata	tatatacaca	6180
cacacatata	tagtgggatt	ggtggaccca	agtgtatata	ttttaattct	taattctaca	6240
tctagcnnnt	tatttactct	gagacatatt	cttgctctgt	tgcccaggct	ggactcaaac	6300
tcaaaattct	cctgcctcctg	tctctggagt	gatggatcac	agttgtatgc	tgccactgtt	6360
tgactcttaa	ctgtgtagtt	gttaagctgc	ttattggtat	gctgggtgtc	gactgctcat	6420
ttcctggaca	cagtgtgtgt	ataattagct	gtagttcctg	ttgctcttta	atcctggtag	6480
cttaattcta	acttgctatt	tttccgtcag	agaaggcaca	agactagttt	ccagtataga	6540
actgtattta	cttccaatca	ggtcaaatat	atatatatatt	gtttgtttgt	ttgtttgttt	6600
tgttttgttt	tgttttgttt	tttttgagac	agggtttctc	tgtatagccc	tggctgtcct	6660
ggaactcact	ctgtagacca	gactgacctc	gaactcagaa	atccacctgc	ctctgcctcc	6720
tgagtgtctg	gattaaaggc	gtgcgccacc	atgcccggcg	ggtcaaatat	ttaaaactac	6780
ctctttttga	atgattattg	atgaattgtg	aattatgctg	ctagtctctg	ggctccctca	6840
tgagggctct	cccatgagg	cacctcacag	gtacagggcc	cagccaagag	gccaaatgat	6900
tgccctaaagt	gtgctttcat	tctccaaata	attcacctca	tccaaaatcc	cacttccctat	6960
attttctaaa	ccacggccac	gttgtgtgca	tccgttctat	ggttcgtttg	attgctttga	7020
caagctcagg	acctagcatc	ccaaagctaa	cagaacccaa	ccgttaggat	attctgaatg	7080
gaaggtcact	tgtgctgatg	gcttttattt	tctccacc	cccatttggt	tgggattgta	7140
gctcttgata	ccgcaccccc	aacacacaca	cagacacccg	gacctatagct	aacacagagg	7200
taaaggagct	gaaacacgtt	tttctctgtg	tgacacactg	gaaaactgat	gaagctccaa	7260
agcttgatag	ggattttgat	atcagatgta	ttatcctggg	tcattgtcgt	gagactggct	7320
cagactctgc	atttgacttt	gagtcgaacc	acagagtggc	ccagagtgc	tctctgcttc	7380
tagccagcct	gtctgacaca	ctggctggcc	ctcctcacca	caatctgacc	tcatgtgggt	7440
gaggggaatt	tttacctggc	tggcccccag	gacagggcat	gggccatggc	agcatcctgt	7500
cgcagttctg	ttgttcagag	gtggatccca	ctgcaggaaa	ctagagtctc	ccatcaatgt	7560

ctttcttctc	agttttatgg	aaataacctt	tccttaatgg	aactgtaatc	ctacaaggac	7620
agggtgggacc	atgtactctt	gctctcctgt	gggtctcaca	gtaacagggc	ggctgaaggc	7680
atggctgtgg	tfgcctttgc	cttccttttag	agcaggtctc	caggaaagca	ctgtgagtaa	7740
gcagcaggta	acagttccta	cttggtgttt	gagatctgaa	gataacgtgg	caagcaagag	7800
gctaagccag	ttcctttctt	agtttagaga	gacatttctt	gactctgcct	ccaggtgacc	7860
tcctctaggt	ctgataacaa	ccagcctttg	gggtttgaga	acattcttgt	tttttgtttt	7920
tggcaacttg	aaaaaagtac	ccagagcctt	tgcattgtta	gcaagaactc	ttctgcctag	7980
acaagcaact	gtgcaacaca	cacagacaca	agcaagcagc	tcaacctaa	catgtgcct	8040
tccaaatgcc	tttaagaagc	ccacaagttg	ggaacatacc	agaggaaatga	ctgaactgcg	8100
ttagaattgc	atgttagtct	acgtgatgga	agaaggcatg	gagtcctctc	taataaaact	8160
gctcatggaa	gactttgcat	gaattccagg	cttccagagg	gttcaaacca	gcaacccaag	8220
gtggatttgg	ttcctgggga	tgacatattt	aggtaagag	agcttaagaa	tcagtctctg	8280
aaagtgtaat	ttacagaaag	tgttctggga	ctgctgttgg	ttcagcctgt	tcactctcac	8340
tctgttcaact	ctggcctgat	gctcatgaag	agttgacctt	taccttagtc	tcagtgtctg	8400
cttttacttg	gtatcctgct	gtgtctacat	ccttgcctgt	tctaacttgt	ctgcttccaa	8460
cactgttgc	ctttaaatt	catttctttt	catttctgt	ctgatagagt	atatggttca	8520
gtagatggct	gtaaagctga	gggctctgca	ttctgggaga	ccctgctcat	gtgtgctgct	8580
tctgtgtgtt	ctctgccagt	gggcagagca	cccaggctct	ctgagcctgc	tggttgggtc	8640
tgaatgat	ctattgatag	aagcttnang	nnnnnnnnna	nnnnntncnn	nnnaaaancc	8700
agtgcagtgc	ctggcatatg	gggtgacattt	agcagtactg	tgaatgatgg	tggggatagc	8760
ccagtgcctgg	attcatgggc	ataaaacctt	atgtgtcatc	ccataaacct	aagaatttgt	8820
gagacacgac	tgtacagacg	aggaaagtaa	ggtttatttg	cctgttctcc	tctctcaag	8880
ttactgaaaa	catggcttat	ggctagataa	gcctcatggc	tagatcactg	ctaaagagtc	8940
ctgacaacca	aattaaatta	ttcttgccaa	aacacaaaat	acagatatcc	atgtgaatgt	9000
aatacattcc	cttacctatt	ttaatccagc	gttatctcgg	aagcagtgtg	cataagtaga	9060
gnacagtatt	gaatctatat	tgtattcttt	gactgtactt	attttttttt	atttgggata	9120
aggtcttacc	atataacccc	agctggcctt	gaacttacca	tatagaccag	tctggccttg	9180
aactcacaga	gatttgccctg	cctcagcctc	ctccaactct	gagattaaag	gccaaacttg	9240
ccaccattcc	tggcttgtaa	gttacttctt	aagtggtgtt	actaaaattt	ttaaatttaa	9300
ggttatgggtg	taggatagtt	tatcttggat	gcattgacata	tttaaaaaata	tttatatatt	9360
ctcttaatag	tttttagaag	gtataaccgga	ctccaaaata	taactgtcta	tttaaatata	9420
aaaacagatt	tattgttcca	taagcttaaa	aattaaagtc	tcaggaaaag	ataccaaact	9480
tggctttttac	ctatttatct	cattttatcc	cagaatattc	actggccagc	aaactctgta	9540
gagcagtgat	tctcaacctg	tgggtcacaa	cctatcctgc	ttatcagata	gttacattat	9600
gaattgtaac	agcagcaaaa	tcacagttac	gcaatatcaa	caaaataatt	ttatggttga	9660
gggtcaccat	aacgtgagga	actgcattaa	agggtcacag	atttaggcag	gttgagagct	9720
atagccacat	agagccttta	caggggtcat	tctcgttgtt	tctatagaaa	acgtttatat	9780
tagataactt	tctcacagac	ttgggttatat	ttccaagaga	tagctgtttt	ataatcccta	9840
ctctaaaaca	attaagattt	ttctagaaag	ttgattattc	acgtgtaaa	agataaaatt	9900
ctaggatatt	tcatttgtat	atgcattatg	aaaaaaaattt	aaatggctca	gaattatgcg	9960
atagctgtgg	aaaagtgcct	cattttaaca	cactttgaac	tccaggcttt	atactgcagt	10020
ttgtttgttg	ttcctcccc	gccccatccc	caactttctt	tcattgctagg	acagaaccca	10080
gggcatgca	cgtgattaga	atatactcct	tcactgagct	gcaccccccc	ccccccccag	10140
ttcttttatt	ttatttttta	ttttgcgaca	ttgtctcact	aagttacctg	ggtaggcctt	10200
gaactcatga	tccttctgct	tcagtctccg	aagtagctgg	gatttagaggc	ctgtgctgtc	10260
agcctggatg	taagagtttg	ttgttgattt	aaattagata	ttgtctcctc	taattaaactc	10320
catttgttgt	cattttcatg	gccttgagta	cacttcaaca	gcattccctgt	tcatactcctt	10380
gaattcttct	ctaaatccta	gcagaccttt	cctgatcttt	cattttctgt	tcacatggaa	10440
gggtacctgca	gccattttat	ctcaagctct	caaatattct	gcttctcagg	acactttctt	10500
atttctttgt	atttcaccat	agaagttttg	catgacctca	ggcatataag	aacaatataa	10560
tcaaacactg	actgataata	aagagtggag	agatttttat	atttttttgt	tttttttgtt	10620
ttgtttgttt	gtttgttgtt	ttgtgttgtt	gggtgtgttg	tttttcgaaa	cagggtttct	10680
ctgtgtagcc	ctggctatcc	tggaaactcac	tctgtagacc	atgctgggtc	cgagtggagt	10740
ttatttttaa	atacatagaa	tttttagcagt	tattaaagat	aaaaggcagt	ctacatactg	10800
tgagatggat	aggtttgtat	agaagaactt	gactttggct	gaatatttga	tactatagga	10860
tgtagagcat	ttccttgttt	ttcagaattc	atcaggattt	tgattttgtg	gatgccagtg	10920
ctaagaatgt	tgtttccaca	aacacattgt	acaatatggc	agaattgtgt	ttagtgtcat	10980
ttcagaaatt	gtttgaaact	ccatttcta	tctaggtcaa	aattcatttc	atggaactca	11040
agccagtttt	tataaatcaa	gcatttcta	gtaatacaat	caaaagtgc	ctagttttgt	11100
acttcatgc	taaggaatgg	cactgatgaa	atattcacct	actttctgta	acagcagaaa	11160
gctctatgta	tacgaaatgt	acttctactta	atggcacggg	attacatata	tgctagcatg	11220

tgacgtgaga	agcacgcatg	ttgcatactc	aaaacagaag	acgcaggggc	agctgcacaa	11280
ggcagcgggtg	ggcagagaca	cttattcatt	catatgtatg	tgggttttgaa	taaagttttg	11340
ctcattgctt	atttaaaaaac	ttttattcac	aatttttttg	accactaaaa	tcagtttgca	11400
acccacagtt	tcaaaagctg	caaaataaga	cctacataac	tacctcgac	aattgtaaat	11460
cacacgagac	ccttggttgg	gtattgtaag	aattgaacac	tgtatccagg	aagtcattag	11520
taaaaaccta	atgtgggtgcc	ttgcttttta	aagattttatt	tactttattta	atatatatgc	11580
cctatcatat	atatacctgc	atgctagaag	agggcatcag	atgtcagtag	agatggttgt	11640
gagccacgat	gtgggttgctg	ggaattgaac	gcaggacctc	tggaagagca	gccagtgctc	11700
ttaaccactg	agccatctct	ccagctctta	atgtagtgc	ttttgtcaga	cattgtgtat	11760
atgggatatt	tagctacagt	tgtttcactt	gccatttttt	tccttataat	tttccattct	11820
tcatttaaaa	agaaatatct	cttatttttt	ttacctgtaa	ttaaatatata	ttcaacagtt	11880
athtagtatt	tgggtgttgg	gttacttagt	attttgtagc	ttttaaacat	ttgttctttc	11940
ttttcctgag	tatgttttgg	tccctgcata	tatgcctgtg	caactgtgat	gtgcctgggtg	12000
cacttggggc	tcatggaggg	cctcatatcc	gttgggaactg	gagtttagagg	cagagctggc	12060
ttatgggtac	cagacctggg	ccctctgcga	aagcagcaac	tgaatcctta	accactgaac	12120
aaaatctctt	cagccccatg	tattttgtac	ctttgtgttt	tatccttgaa	ataaatggcc	12180
ttttaagaaa	tgagaaaagc	ctttaatccc	agcagaggta	agtggatcac	tgagttcaag	12240
gccagcttgt	ccacatagtt	ccaggagagc	cagggtctaca	cagaggaaaa	aaaaaatnca	12300
aaaaacagga	aaaacacaca	cctccttgat	tttaaggggtt	tttgtttgtt	ggttgttttt	12360
tttttttgag	ttttggggag	gggggtatatt	ttttaatgtg	tctgtagtgtg	gctttgtttt	12420
aagcatttta	atcatacttt	atttttaaaa	aaactaaaag	cttttttaag	gctaggtctt	12480
gctatgtggc	cctagtgttc	ctgggacttg	ctctgtacac	cgggttgact	ctgagcctgt	12540
gcgccttctg	cctctgcctc	catagttaga	ttctcaggac	atgttacaaa	gactgtgctg	12600
tgaagatgag	tttttgttcc	tgggagggaa	ggttggagct	gacttgtgag	gtactgactt	12660
gggtctgcct	tacagttgac	tggattgttg	cggacatgct	ggctgccaga	caggatgcc	12720
taggacatgt	gcgtacgta	ctgaaagaca	agttaaaatg	gcttccgctg	tatgggttct	12780
actttgtctc	ggtaaaacttt	gtctttgccc	ttttatttca	aacttaaac	catttaata	12840
aactatatct	gatttttttg	tttatgtgtt	tgttttatgg	taccctgtat	tgaacatggg	12900
gtcataatgt	tgctactgag	tgacagcctt	agttcagaca	tttttttaag	cgacttttac	12960
tagtattttt	atttagaatt	ctatatgtgt	gcacatgcat	atgtgtgctt	gtgtgcacac	13020
gtggatgcat	gtgaggtcga	aggacaattt	tcagtacaag	tgtgagtgct	acttttttag	13080
caccttccac	tcttattttg	agacagcttc	ctagaccttt	gctgagttgc	ccaggctagc	13140
cggccagtga	gccctgggca	tctaccgttc	tctgcctcct	tacctttact	taggttacaa	13200
gtgtgtgctg	ctacgcccag	ctgtttacta	gattctaggg	atccaaatgt	gggtcctcgt	13260
aacttgtgag	acaagtactt	tccaaactga	gccacctccc	tagctcttct	tcacggttcc	13320
tgtatgtgtg	tgctagatg	gctggttgtc	cgtatattta	agtccagtag	cagaaataga	13380
aatacctagg	agtccaatag	aaagctacaa	gtgcagaatt	gacaatcggt	aatgttcgga	13440
aattgattca	aaagtagtta	gtgagtgaca	gacaggagct	aaaagcagac	tctgagctca	13500
gagtgtgaag	tgtggagaaa	tgtgttttct	cacagttctg	aaggctgaaa	gtctccccc	13560
ggtcaggatg	tgggtggtac	tgctgtctcc	caacacccac	ctctttggat	tatagactgc	13620
agccttctcc	ctgtgttctg	agccggcctt	tcccacatgt	ggacatcctt	ggtgggtgtt	13680
ccaccagcag	ggcctcagct	agtgccctta	tttcaactaa	ctgtaatgat	tttcttaaa	13740
accctgtctc	catcacacagt	cactgtggaa	gctgaagctt	caatgtaaga	gttaaggggg	13800
gagggggaaa	tttagtccat	aatgggtgtca	caccaatctc	tgtagctgag	tccatgattc	13860
agttctttta	aggctctgag	tgtagacatt	atcttaatta	ttttgccc	ttatgtatta	13920
tcttttaatt	attttatgta	actgaatgcc	tgtgtatata	tgtttctggt	tcctagtcca	13980
tatttttaatt	ccttaaagga	tggaggtgta	gacttttgct	tttttaattt	tctatccttc	14040
cctcctggcc	tcctgtggcc	tcttttacgt	attttatttt	tttaatttat	tttatgtgtt	14100
tgagtgtttt	gtatctatgc	attcctgggg	cccatggagg	tcagaaaaac	acattaggtg	14160
gcctacaact	gagttatggg	tgggtttgtg	ccatgggggtg	ctgggacttg	atcaccagtc	14220
tctgagaaga	ctcgtgtctg	ctgagccttc	tctccagtc	tgggagtggtg	gatattttta	14280
ggataactttt	aattgacttg	gtgaatgaca	gtagaaaatc	aatgagttag	gatccatcgg	14340
aaaaagcttt	tgaactaaat	cttttaagaa	gaaaatattt	taagtgtctaa	caaaattaaa	14400
tgtgtatttt	ccatgatgca	gttttacttg	ggctctgtag	aaataggatt	ttcagggtaca	14460
tattgtatat	atagttggca	atatttaaat	actaactgtc	gcttgagttc	tgaaatgtag	14520
ttttatgttt	tttactcatt	aggagtagag	ttgccttaat	aactacggag	attagttatt	14580
aaagaataat	tgctcttctt	ttttcttttc	tgtgtaccag	catggaggaa	tttatgtaaa	14640
acgaagtgcc	aaatttaaat	ataaagaaat	gagaagcaag	ctgcagagct	atgtgaacgc	14700
aggaacaccg	gtaagtgcgc	ccgcttttat	tcctcaaggc	aggtttaagaa	gttaagttct	14760
taagtcattt	tgaataatata	ttaccccatg	tggagcaatg	gaactgggtc	gggggtttgt	14820
tgagataagc	tgctcctctg	ccgtgaggta	agattgctgc	agggtattgt	aaggtttctc	14880

ctgagtaaca	gtcagcatgg	gctcgggacg	ggcaagggca	ggccttagtg	tcgagaggat	14940
ggagctcact	gaagcccaaa	agagttagtc	ttcacatgag	attcagttct	agaagaagtt	15000
aaattgcttt	ctttctgtgt	aaatttggat	ttttattgta	gaaattaaag	tttgttttct	15060
tttaaaacaa	acacaaaccc	agagcaaaga	gtctcctagt	gaagagtcac	tccgtgtcag	15120
tattttacac	aaptgttttt	ctgtaaaggg	ggaaaaagaa	ttcaaatctt	ctctttcaag	15180
aatgctgact	gctgccaact	gcctctcccc	gtggcccctc	tctgtataga	caggcatagc	15240
tatggtgagg	acttggggcg	ctcttgtctt	tctcctctct	ctgcttctct	accctttctc	15300
tcgtgcccct	cacttaccag	gccctgggaa	gctacacacc	aggcaacagt	gaccagggcc	15360
tcggcctggg	cttcgaccaa	ttactagagc	agaacacagc	gcagctgcag	tgttgttttg	15420
tgctgtgcac	tgtattaggt	tgtgttttca	tcacctttgg	gtttgtgat	gttttgatga	15480
agtcctggta	ccattctagt	ttttacattc	tgggtagata	gagtttattc	aaggctctca	15540
ggcatatgaa	tggagagact	cctctttaca	gccattcgtg	tagcatgcat	aactgctctt	15600
ctgtattctc	tctagtgtct	ttttttttgt	gtgtgaatct	gatgtcttgt	tattcaccta	15660
caatgtggag	taatggtcac	aaacatataa	agtacttatg	cctttatctg	ccaaattgta	15720
tttaactttt	cagcttttaa	tataactttt	tatataataa	ttaatattat	ttaaaaaaa	15780
ttgaatacca	gcctgtttata	gtggcatatg	cctgtgttcc	tagcactcag	gagacaaagg	15840
cagaagtgtg	agaacttcag	actcatactc	agctatatac	aagaccccaa	atttgtgcta	15900
gattctgcag	tacagccatg	agtgccccca	tcttagaggg	agatcgctca	tccttgtgct	15960
gttctttaag	tcttaccctg	caaccacact	taagtacact	cttgctcaca	gtcctttaga	16020
atctcacact	ctttctcttt	acagacacca	tgtcattgcc	cactttatta	tttatctgat	16080
gtctacaaa	attatgaaag	agaaaacttg	atgcattctg	tgtaaagtag	ttgacacaaa	16140
taatagtatt	caagaatgac	ttcttaaatg	aacactgaat	gaatagtttg	ttctaatttt	16200
tttgatcaac	aaatcaaaaa	atatttagat	taaatatcta	agatacaaa	cataatacca	16260
catgaatcat	taaagttagt	aatcaatctt	ataagtgaat	gaccctaaaa	ctcatagaca	16320
ttaataattg	ctttcattgc	ttagatataa	actttattga	ttaatacgtt	ctcatgaaag	16380
tgggtcttgg	aagggtcttg	aaacgaaaat	atttttctta	ctgctttttt	cttctagtaa	16440
ctgattgaat	ttttctgcag	ttccataaag	catctgggtc	attgctatta	tccaatatga	16500
ggatatataa	caaagtattg	atttttaaat	ttggcgggtg	taagacaaag	ctggcgtgt	16560
gaatgagggg	gtctctgttt	cttgccccct	ctcttgggtt	cttttccttt	tgttgggttg	16620
ccttctccag	ctgctatgtg	atgggttctg	attatcttat	tatatcttat	tttgttattt	16680
ttcattgtta	tctcttagaa	gccaacatgt	tatatcacct	ccactcccac	cattagggtg	16740
ctcacaaaata	ccccaaagta	aacaaccaca	tcattgtcat	ttctgtatag	ttccataagt	16800
gttggttaact	ctactgactc	ttgtgagcag	gcccaatttg	ctttatccct	agctgggtga	16860
cctgggttcc	tcccaaacac	cataccgtcc	atcaaaactg	gtccttttcc	aagcacacac	16920
cagatactgc	tcactctgag	actcttctca	tccacctaag	gactgcctgc	tccctggcag	16980
aaagggcctc	tagtcccata	cccttaccgc	ctcaccaatg	ccttaggaac	atgtgtcaca	17040
tgccctctgt	gggtcatttcc	gtttacagta	gggaaatttg	cctgataact	tcgagcacac	17100
ctataaagag	gccttgcctg	ctctcatatt	tagctggaga	agataatgta	ctcaccaact	17160
ccactctatg	caaccagtc	tgtcttgccc	atgccagtc	gacgtgaatc	ttacacctgg	17220
attcagattg	atgaatctac	aacatcaccc	actccatgct	tccttctaaa	tcagcagttc	17280
tagcctgaat	gacagatgct	acccaagtct	catctagtta	gccctgtccg	gagtaacctt	17340
gaccttgagg	attagaccag	gatgcacatc	ctgcaccagt	tccctttgtc	cacctgactt	17400
catcccacc	gggcccatag	ccatgctcag	gctccaccct	ccatgcacaa	agctggcttt	17460
tccagcttcc	ttcacctgta	tcagacacaa	atagcaaaag	gggtccacgt	gcctagggtc	17520
catcacaaaga	ccatgtgcgg	tagtttggaa	aacagtctcc	acttgaggct	cagatagmtt	17580
ggaatcttgg	ctctcatgta	gttgtagtga	ttagatcagt	ttaggaaagta	tgacctttnt	17640
ggaaagaaca	tataactggg	aggggctttg	agatgtaaag	gccccacaca	attcctagtt	17700
caaactntac	tgccctgtca	aagcttgagg	catgaactct	cactgttccct	gatgtcatgg	17760
ttcctgtctg	cttcacaaat	tccttatcat	taggggtccc	tttcttccct	ggaattttta	17820
gataaataaa	cacttctttt	aaaacaacaa	gaacaacaaa	tctgacnctg	ataatggatt	17880
ttaaggcgtc	ttctctggat	aagaaaaaaa	aaagaatata	tttgcatagg	tgctgtatta	17940
cttttgtcat	tgggtataacc	tgactggaag	caacttaaa	gaagaagaat	gtatcttgat	18000
ttgtagatta	agagcaccat	gactaagaag	gcatagcagc	acagggtgcac	cagcaagaac	18060
ataggctgct	agcttcagatc	tctgtagata	tgggaacagg	gcaggaaagct	agtagtctat	18120
aaacctcagg	acccatccca	tggagttcct	tgtcttccag	tgatgtccct	tgtcttaaa	18180
tttcacagtt	ccpacagcag	cacctgccgt	ctgggaacca	acctgtgggtg	gatattttac	18240
aacgtgatag	gcataattttg	tctctagccc	tgtaggttta	tagccatcct	atacttcagt	18300
tatctagtc	caactcagtc	tgtgtgtctt	atagttccaa	cacttcaaaa	ctacaaagtc	18360
ttaagggccca	tgggctcggg	tttattagag	cagtaacacc	tctactagct	ttctgtgtta	18420
ccactcctc	ttagggtctg	gttgaaatcc	taataggaag	cagcttgaga	ggagggttta	18480
ttgtggccca	tactttgttg	gtacattcta	tcattgcaag	gtggcactgt	gatacagccg	18540

agggcatccg	aggatgggtac	tgttgggctta	catctggggtg	ggacaggaaa	tggttaattct	18600
caaggccccc	ctgcttgggtg	acttcttttca	gttaagcccc	atactctaaa	tcctctacaa	18660
cctcccaaca	taatgccacc	agctgggggat	cagctgttga	cagtgtctggc	ccagggggagc	18720
agttaaatac	cagaccaggg	gacctgaaaa	cagagaactg	cagaggggct	gtggggacttt	18780
ataccagctt	tgagacaaa	tcacggcatt	tcttgtgtgag	cttgggttcat	aaacaaatat	18840
atattctcct	ataggctcct	ttagtgggtg	tttcatatcc	acaaatttgt	tcagaaaaac	18900
actgtgtttt	atgctagctg	tgtaggagat	aataccgctg	ggagtccactt	gagcatggat	18960
aagtgcata	gttcgtcctc	atgagtcctt	gtcctgtttc	tgtattatgt	ttacttgatg	19020
agtttagttt	gtcagttggc	caccaattaa	aaagtatcat	tttatttttt	ttacaatact	19080
cagttctcaa	gttaggagtt	ttgttattat	atggcttcaa	tattcacatt	ttaacctttc	19140
caggagttaa	gtataaaaa	ttatatcaac	tgttgactta	gtaaatatct	attacagata	19200
ctatattctt	cttagtttat	atcatgaata	tgaggttgct	taaagtaagt	gatgtaaaat	19260
acactagggg	atgcttataa	aatggaatgt	tgtgagtttt	ttgaaacacg	agtactaaaat	19320
tcataagttt	ttaaatagtt	acactgttag	cttcagtact	gctagatata	tgtctataat	19380
ggctgaagag	tggagcttgg	atattataag	tgtactctgt	atattcatgc	agacatatag	19440
cagattccac	tagtatgtgt	ggttaatatg	tgctaataaa	aatttaatac	aaaagtcctg	19500
ttttattact	gggaaccaga	gggggttgggt	gtgctgattt	taagtcagtg	actattagca	19560
tattctaaga	aacagtttta	ggattttaaa	gattggcttt	accataaatg	tagagctatg	19620
ttttactata	atccatatta	tggctggcct	taattcaatc	tctgcagttt	ggttactctg	19680
ctcaaagtga	aggctattta	taaatgatac	acattttctc	accataggaa	atactactg	19740
gccaataaca	gagttagaat	tgctaattg	atgggtacca	caatggactc	aacacaaact	19800
aaagtttatt	tatgcccaca	gatgtatctt	gtgattttcc	cagaggggac	aaggtataat	19860
gcaacataca	caaaactcct	ttcagccagt	caggcatttg	ctgctcagcg	gggtaagtaa	19920
agatttaact	gtattcagaa	aaacactttt	ttaagaagag	tgatctttgt	ttccttcaga	19980
gtcataacta	agaatatgcg	tttcttgtta	gagctaaagt	agagaatatc	cgatcttcta	20040
cagagttagg	tatattctta	ttagtctgtg	tctgagaggt	tagagacgca	ggcttgctat	20100
ggcacatttc	ccatgctgtg	aattgagtta	aaaatgtagg	taaatgatat	ccccaaagaa	20160
gtatactttt	ggagtgtact	agtataaagc	ctgggtgttat	aacataaaca	cgcacgtgcg	20220
gatgtatatg	tagcacatat	gtaaacacag	gtatatgcat	tgtaataaga	aagtggaggt	20280
cggggccccc	tgccagcaag	tcttttagtg	atgctgagct	aatgctgaga	ggtagaaaga	20340
ccaagaaggg	tggagttgct	cattcgccaa	aggctcagagc	tcactgtgtg	ccataactcg	20400
agtgttctgt	ctcccttttg	atacagtttt	cttgttttta	attattagtt	tttacaatta	20460
tcccataaaa	tgtgggctca	ttgtgggtcat	cgttttcata	aagtccttca	agtatacacc	20520
cagcaagtat	ctaaatacac	tgggaagaat	cagtcagctg	atggcttgaa	gttttcaggac	20580
atctagtggc	acatcatgct	tcagaaccga	cctgcactta	gtcagggtca	tattcatgcc	20640
acgtgaagag	gagaggaggc	catgccgtct	gacttaggat	ggaaatttcc	ttcgagcaaa	20700
cacgaacggg	ctaggtctta	gttataggca	tagtgtctgt	ggttatacta	ggcagacatt	20760
agtggactgg	gtgttagaag	gtacagacag	gcaagaattt	gctgtagatt	tgttttccctc	20820
atgtgttgac	accacatcta	acctgctttt	tgagcttcta	gtcctaataa	tctcataaaa	20880
atactggttg	aaccagaaat	ggtgttgcaa	agctatgata	ccagctctcg	ggatctaggg	20940
tgggaggatc	ataaatttga	ggccagcttg	ggtctgtctt	agagaaaaaa	gaaaaataaa	21000
aagtctggtc	aaggtaacat	ggagcctgga	agtttcacag	ggtgattctg	taaaggctcct	21060
gagacaagat	ggcctctagt	ggcgaatgac	ttagctgaca	agaaaaactt	cccagcttgg	21120
ttgacttttc	agacttcata	caagtttgtg	aataaattac	actccttctg	cccttggggac	21180
tgaactcaga	tatgtggttg	tgggaatggc	tttctttccc	acaccacctt	gcattttaaa	21240
aattcttctg	tagacagtcc	caccatcctg	tagctgttct	tccttatgtc	gccacttttc	21300
ctggagagag	gcagtgcaga	cttcaaccgc	cttctcccta	gtcgtgttcc	atagcacatc	21360
gaaagaccta	gtgcttccctg	tgaaattgta	agtacatcct	ggagtccagg	agaggaggaa	21420
gccgaacaga	gtggagggaa	tgctgagttc	tgtcctaaga	aagactgcgt	gcttagcaag	21480
atgctgtctg	tctcctgtcg	tgtctttctt	gtcagaactt	atcaaagaga	aggctcgcag	21540
tgggtcataa	tcttcccaag	gaccagcctt	cccagcttct	cgcagcatat	ctcattcatg	21600
tagatgttta	atggatatgt	gtcaatgggg	ttgacctaa	tgagatggca	atgtatgtga	21660
gcattctagg	tgtgaggtta	tggcattaaa	ctttaatttc	cgtctatttg	tggttagttga	21720
taagtaattt	agatgttgac	tttcatgtat	tcctaattat	gaccacattg	aatctacctg	21780
ctttctaggc	cttgagttat	taaaacacgt	actgacacca	agaataaagg	ccactcacgt	21840
tgcttttgat	tctatgaaga	gtcattttaga	tgcaatttat	gatgtcacag	tggtttatga	21900
agggaaatgag	aaagggttcag	gaaaatactc	aaatccacca	tccatgactg	gtaagtcctg	21960
atttccatag	aagtgaata	gtacatggta	caggtaagat	aaactcttgt	tggttcgctt	22020
tgcttagctt	gggttcagttt	gggtttcagt	agagggttcc	actatgaagc	tctggctggc	22080
cgggaactca	ctatgtagac	caagctggcc	ttgggctcca	ctacaccag	caccaatcac	22140
ccactcttat	cttttatgct	ttttgttttt	gctttgagct	ttctttataa	catgtttggg	22200

aaggacattg	tcattattta	caagaagaaa	tatggtcttt	tcccaacatg	ctagaattta	22260
aagactcaga	actcttgctt	ttgtcagtg	caaagtgaga	atggctgtga	agtgcactgg	22320
ctttgagtga	gaatagttca	ggtaactata	gccacagact	caacatttga	acatgggaac	22380
agggtgagaac	ggagtgtatg	aagattcttg	cccctttcag	agaattcatt	ttagagagag	22440
atgagagttag	taaggaagag	agaagagaga	gacgtggtat	tttgctgcag	actaaagaga	22500
tctcttataa	tcgcagtagt	aaggagggaag	aagcagaaga	tgatgactac	agggccaggc	22560
tgaacaatct	agtaaaatcc	taagtcagga	agtcagggct	gaggtgcagc	tcagtaggag	22620
tgattgtgtc	tgccttacac	aaggcctgga	tttagctccc	agtagcaacg	aagggaggcg	22680
aggggtgggca	aaatcgaaca	cttactcttg	gagactccct	ttatgaatat	taccacactc	22740
cagtaaatatc	tctccagaga	tttcagatga	gattctgctt	cctggtaaac	aggaggccaa	22800
gaatattatg	tcacactgaa	catgggatgg	aagacatgtt	ctgaggaatg	tctgcactcc	22860
agtgtgatga	agacttgaag	tttagggaca	ttttccctcc	ctggccccac	tcaccccatc	22920
tgtattgagt	attccccatg	tgctcatctt	tatttgtatg	ttactttcca	ggaaggggaa	22980
gcagattgat	atccaacccc	agccagtttt	cttaaatact	ttgtggatgg	gattggcttt	23040
gacagtaaat	gaggaaatgt	aaaatgtaaa	agattctaat	ttttaatttt	ttaaagggtga	23100
ggttttctgt	tagtacgcag	agttagaggt	ttcttactga	tgcttgcgta	ccttagaggaa	23160
ggatggctac	ttctccaagg	cttgctgtta	gaagtcagtg	acatgggctt	aacaagagat	23220
atgtgctaata	gaggttttaa	tttcagctta	atactgcaaa	tcataagtcg	atagctttta	23280
tgtttttaaat	tcttttagtc	ttaatgtttc	atttttacca	taagttactt	tgtataatca	23340
caaatcttaa	actagtaaga	cgtgaaattt	tcttcttctt	tgtagagatt	tctctgcaaa	23400
cagtgcctcaa	aacttccat	tcactttgat	cgtatagaca	gaaatgaagt	tccagaggaa	23460
caagaacaca	tgaaaaagtg	gcttcattgag	cgctttgaga	taaaagatag	gtaagtgtta	23520
agagctccag	catttagaaa	gtgcagttca	accaaatttt	actctcagat	cctgcttgaa	23580
aggagtcttt	ttatcttcat	tatttagtaa	atactaatac	tacctgcata	gacaagacca	23640
catatactta	aatgtagcat	gtttcatggg	gcgttaccct	tgtttaacaa	ttaagtttaa	23700
catcctacat	cagtttgctt	gttgattctt	gtaccatgac	aactcaaac	acgcatgcgt	23760
ttattccaaa	gtcgatagca	cagcaaaagt	gaaactaaag	tctgtattgt	ttcaagaatg	23820
ctttttgtga	actcgggtta	aatcttattc	tatctttctg	tgttcacatt	gtacattttc	23880
atgagtcact	ataaaaatca	tgacatgggt	gcctacctgc	agtgtttgtc	ggacagtagg	23940
ctgctgtgtg	ataagagcct	ttcctcttca	gctacacggg	ggacacgagg	ctttgggggt	24000
caagactgaa	gcacgggtga	gcacaacacc	tttgtgttgt	gggaagggaag	ggaattgttc	24060
ttttcataat	gaaattgtcc	cccttcttga	gttagtagaa	agtattacaa	ggatagagag	24120
ttgaaatgaa	gctttatatt	agatttatgc	cctgtgttgt	cacgtgtttc	tacctgacat	24180
aacttttcaa	cccagccgct	caggattatt	ttgatgtatg	gaacaatgta	agaaggccta	24240
tgatcgggta	actcactgtt	gtagctctgt	ggaagcggct	cacagggcag	agggacgctt	24300
ctgtgctttt	gtgcctgtcc	tgctgttaga	atcttacaga	ggaggatgaa	tgaatgaccc	24360
tttttatttc	tcttgctctg	ttttcttaatt	ttatgggaat	aagaactttt	ggtagggtct	24420
tgctactggc	ctcttggtgt	gaagagacac	cttgagcaaa	gcaactcttc	tgagagaaag	24480
catttagttg	gggaattcct	tacagcttca	gaggttgagt	tcgttttcat	catgctgagg	24540
acagggaggc	actcaggcag	gagaagtagt	tgagagccac	attctgatct	acaggcagag	24600
agagacagac	tgagcctggc	atgggttctt	ggaacctcaa	agcctctcat	ccctacccca	24660
tctcccgacc	cctatacaca	cttctctcaa	caaggctaca	ccttctaata	cttcttaaag	24720
agtcaccaca	tccagcgact	aagcattcag	atatgtgaac	ctgtcggagc	ctttcttact	24780
cagatcacct	caggaggaaa	actcctatgc	tataagaatt	tcttttcttt	cgcattctttg	24840
aaagcttggt	tttgtgtgat	tagatcctgg	cctcacacat	gctcggcaat	catctttactg	24900
ttgagctcca	gcctcagccg	ttttcattgg	cttatgggat	gcgagccatg	ggagagaagc	24960
tagaaggcct	ttcgttttat	gagtcgggtt	ggtggaacca	cttacagatg	gaagattttac	25020
aaacaaaaat	gaagctgggg	ccatcaaggc	tcagcactcg	ctgctctttc	agagagttca	25080
ggttcatttc	tcagtaacca	catggtggct	ttgtaaatgt	aacttcatat	tcaatgaccc	25140
tgacacccct	ttctggcctc	tgtgggcacc	agacacaatc	atggtatata	gacacacaca	25200
ctagccaaca	cccattctaca	taaaagtata	taaacatatc	tttatcttaa	aaatccccga	25260
agtcctcatt	aaatatctta	gatccccgcc	gtgttttgat	ttttgtttcc	cacgtgggtga	25320
ggatataata	tcattgtccaa	actgtaagga	gtgaatgcc	tcccgtgcct	ctcggacacc	25380
tctgcactca	tccaagtttt	ctaaggagct	gtacttgctc	agcaagtact	caatacctaa	25440
taaatgggtt	atgtttgttt	caacacccaa	aatgtccaaa	actgaaagat	caattctgtt	25500
gttttccctc	tggccatagg	ttgctcatag	agttctatga	ttcaccagat	ccagaaagaa	25560
gaaacaaatt	tcctgggaaa	agtgttcatt	ccagactaag	tgtgaagaag	actttacctt	25620
cagtgttgat	cttggggagt	ttgactgcgg	tcattgctgat	gacggagtcc	ggaaggaaac	25680
tgtacatggg	cacctgggtg	tatggaaccc	tccttggtcg	cctgtgggtt	gttattaaag	25740
cataagcaag	tagcaggctg	cagtcacagt	ctcttattga	tggctacaca	ttgtatcaca	25800
aattaaataa	ggagtgttct	tgtgtgtgtt	ttttttgttt	tggtttgttc	25860	

tggtttaagc	cttcatgatt	gaacactgga	taaagtagag	tttgtgacca	cagccaacat	25920
gcatttgatt	tggggcaaac	acatgtggct	tttcaggtgc	tggggttgct	ggagacatgg	25980
aagctaagtg	gagtttatgc	tggttttttt	ttttttttaa	tggttttcag	aattaatgtc	26040
cacttgtaaa	gattattgga	tactttctgt	aattcagaag	gttgattttt	aacactagtt	26100
tgcatgatgt	ttcgctatat	tggttatctt	ccatttgact	acttggcagc	tcagactcct	26160
aatactaaag	tattttacat	tttgaagcta	tgatgactg	gttttttggt	gttggtgttg	26220
ttgttaattt	ctgaaagtca	atgaaagaca	ctgtaatgat	gcgttaagat	gttccaagaa	26280
aaaggtgaga	attattcatg	gcaaaaaaga	tctgtctagt	gtatattttt	atttatattg	26340
tctatttagc	taattttctt	tatatattgca	aaataatgaa	catttttaat	attttattaa	26400
atgcttgatt	tgcatacccc	cgattctaca	gagaataatg	tgtaaaagt	cagaatagac	26460
ttgaagctct	gctgtgactc	agtctccttt	gtcagagctt	ctagtagccc	agctactgag	26520
ctgctttggt	agtacctcca	gcacctgagc	cgtaaatgac	ttataaatgc	aagggaccgc	26580
ttatcttcat	atcggaatag	acatgaacag	agctctaagg	cgatgaaagt	ctgccagcat	26640
cctctctgtc	ctcgacgtg	cctctctgct	ggctccattt	gctttggcac	tcggttcgat	26700
ctagagtgt	ggtgctcact	gcttatttca	gcctgggctc	tggtggtttg	tgctctccag	26760
tggtgctggt	cactgttggt	gtgcaggtgg	tgctgcccgt	actcagaggg	gcagctccct	26820
ggctcctgag	ggtgagcctt	cttggtact	acagaagtat	tggtgctttg	tgatggcaa	26880
gaaccatcag	gattggataa	atgtgttatt	tctctttgat	ttccatggag	ccacactggt	26940
ggtacatgtc	ccctgtgaac	agagctacct	ttcaggagca	catcatactg	tcgtgagtca	27000
cggtcacggtg	tgctctgtga	gaagaggctt	tctaactgtg	gatttgccgt	gtttctatgt	27060
tggtatttaa	gcgtgattgc	ctactagtca	ttcaaggtaa	catttctgca	aatttcatac	27120
agatttttgt	cacaaaatta	ctataccaat	gatctagttg	aaatagacca	attgaatcac	27180
aataaataat	tttttttaat	tgagggaaaa	tttgcttctt	gttttttcaa	agccagaaaa	27240
cgagccattt	caaacatctt	tgaagagtca	tggtctgtca	cttggtttct	atgtgttagt	27300
gtctatatct	atgtatggat	acacatgaac	atgtatatct	atacacacac	gccaatagaa	27360
tataacagcc	taaaaacaat	ccagcttggt	tatcatgtta	ctgtgctgaa	ttgtaatggg	27420
ttttacttac	aaagtgaggg	taaaatcgat	ttcatgtctt	tggttaatac	gtttttttca	27480
gcaatcctat	tagagcttat	tttgaccaga	tcaaaaataag	tacaagtcca	gagactttaa	27540
atatggctga	ggtctagagc	gatagctcag	tagttaggaa	cacatgccac	cttttcaagg	27600
gcttcagttc	ccagcactca	tatggaggct	cacagaaggc	tggaattcca	gcttcattga	27660
attggacaca	tcctctagct	tcctatggat	tgctctgtct	tctctccctt	ctctctctct	27720
ctctctctct	ctctctctct	ctctctctct	ctcacccttt	aaatatcatg	gatatgtctg	27780
gcattttaa	tttaagacac	agaaccattg	gaattacatg	gattatagct	gattctcttt	27840
gaacagggca	cagtgttctg	cgtaagatct	cttgatcatt	agcactggac	tcactctcct	27900
cacaagtagc	ctatcaaatg	tggtattaga	aaatacattg	tgtaaaaatc	tttgaaagat	27960
gagaagaatc	tcctaacaat	gtttattttg	acttgacatc	actatttctt	gaaaattaac	28020
tgctctatgat	tcttttcaca	tagtgttaaga	tcttacttgt	atcaccatca	gcttgcagct	28080
taggggctgc	agttgttctc	cttcataaga	ctgccatccg	tggtgatgct	tttatgtttt	28140
tcagaaagga	tggtgggtg	aaagtaagaa	aacaaagtct	cttcttgtct	ctcatgtctg	28200
tgatcactag	catttcacaa	ctcagggatt	catccatttt	ccagcagata	aaagggttag	28260
cgattaaccc	tgcatcttga	gtttagaaag	ctacaatatt	ttttaaatat	tgagcaatga	28320
ttttaaaaaa	atacattgga	ataccceaaa	ttgtgaagca	atccaaaagt	tggaactgtat	28380
aagctaattt	gcctacttta	aaggatgtga	ccctcaccca	ggaaacctgt	aggattttact	28440
taacaaggct	ttacatgaaa	atgccaccgt	ggccatttct	taaacactgg	tggtctcttc	28500
cagatttcat	ttctatgttt	gtttgtttgt	tggttttttt	ttacttagat	tgctgtgagg	28560
tttttttttt	ataacaaata	tacatttttt	tctttgtcac	attacatgct	ttgtcaatca	28620
aatgacctaa	ctaggttgcc	tattaagaaa	actacatatt	gaaatctgcc	aaaatgtcgg	28680
cataaacaaa	ctggctccta	attgtgtacc	agatctacat	ttgaaagaac	agaaatgtct	28740
cacaagacaa	taaggctcata	tgtaaaacac	taataaaact	ttaacctcaa	caattgtttc	28800
tgaagtgttg	agattaaaga	ctgagtgttt	gcggaacgtt	gacatgtcca	tggtccaggct	28860
agtttctcgt	tttctttttg	tcttaagact	aaacattggc	tggtcttaaaa	tattaccagt	28920
tctatatagt	ttacattata	gacagaatat	ataacattta	agtattagta	tgaaaaatcag	28980
tacttttggtg	agactaatat	ttggaatatc	cagatgattt	gatatcatgt	aggtaaagta	29040
agtattttgtg	tgactgactg	aacttaaaat	ctcttattca	tatatcatgg	ataacagctg	29100
ggagttgtga	cacatggctg	ccatccaggc	actcggaaaa	tcagggtttt	gagaaagaga	29160
gtgtttttcca	agtcagcctg	gtctatatag	caagcttcag	actagccagg	actgtgtacc	29220
aagatcttct	tcacgccacc	cacacaaaca	agagaagtta	tatagagaat	tgcttgggat	29280
ttagttacaa	catttttgtt	aggatttcac	ttaatgggca	gggggtgggg	gagttagcag	29340
tttgcatttt	cagagaatgg	gttccattcc	cagcatccac	agggcagtaa	ctgaagtaac	29400
tgtagttcca	gggtatccaa	caccttcata	tggaacacaca	cgagggcaaa	agaccagcat	29460
gcataccatt	aaaatgaatt	attaaaaatt	ttttaaaaaa	gactttgata	tatttttagt	29520

ttgtgtatgc	tgaggtggat	ctgttgctct	tggcaactg	agtagtaata	cgtggagttt	29580
tataagaggc	cagagatctc	actttcatat	aataacatt	ggataataac	ctagtgtgta	29640
cctaagcaaa	gaagtctctt	ggaagagcta	catcacagaa	aattggtatc	agagcatata	29700
ctcagtcct	ggttatgcca	tgagacaggc	cagtgtttta	cccagactag	tgttttgtca	29760
tatctgaata	aataaataac	aaagtcgtat	tctacaactc	taaactctca	tatagaaaaa	29820
aaaataggaa	tgcggttgac	acaatttggt	gcctgaagtg	atgaggaaat	taagtacaaa	29880
agtaatcaag	aagtactcag	ttactgtgtc	ttcttatgta	gcttcagaat	aagttctgat	29940
agtggttactg	ggtttccctgg	tgctttataa	tccagtacaa	gggaaaaat	gaacctcatt	30000
tttttttata	aaaatttggt	gacaaaaaag	ttacattggt	ttataaagtg	ttttgtggtt	30060
tatgtgtata	tgtgtgtttg	attggcagat	atagctataa	agctacatta	tttcagatgg	30120
gtattttgtt	tgagatttgc	tgggtgttcg	ctgtatgtcg	actgcttctg	cacctcttgt	30180
aggcaccagc	ctgtctgcag	gcagaccacc	caccccatgt	tgggtccttc	tgctctgggt	30240
tagccatcca	accttggaac	gctttaaata	aagagtgtcc	cagggtttatc	acctgtagaa	30300
cattgtctacc	acctgtagaa	cattgtctacg	atgttagcca	caaagaattg	ttgagagcat	30360
gtgaactact	gcagatgatg	cagcagatgc	agagcataga	aagaattgtag	ctggtgtctg	30420
ctcgggttcc	tgtcctctgt	cttcatagat	gaccttcagg	tcagatgggg	tttggataaa	30480
gtctagggat	gggatgccat	ctgggtcctt	cctgtttctc	aaggagacag	tgggtgtgtg	30540
aagtggatgc	tctgtattat	ggttcagatt	gtaggttttg	tttaagattc	agatcacgtg	30600
gtgtagagag	atacaacaag	agaaagaaca	tctccattat	aagtcacatg	gcctctcaac	30660
tacagctcaa	tagaagctgg	atcttcaggg	ctaggagtg	gctcagtcct	caccttttgg	30720
gcacaggtgg	aggggatggc	atggcagctg	cagtaatgtc	tgaggggttc	cttctggcat	30780
tccgcactat	ggtacgtttt	agactaaagt	cgagacaatg	cctatcggag	tgcccttcta	30840
tccgtgtagt	ggggagtgtg	gaagtctgtg	gagaggtcag	tgtggtgact	ccatggagtg	30900
tgggttccca	tatgtaaagt	tgcaactcagc	tgagctctat	ttgtcgaaag	tgtgagggtta	30960
catctgttca	ttattttctg	acatatatta	gagctcagaa	ttgtcaggaa	tgttcatgat	31020
taaacattgt	tccttgtcat	ttggaacct	gcctatattg	gcttctgtcc	tatgcttcca	31080
cgagttaact	ttctgtctgg	gccaggagtc	gttttgattt	ttacaggcat	ctaacatatt	31140
tctgtttgtc	tgataccccc	ttccttctgt	tttcttatct	gcactgtctc	ttcagtctgt	31200
cgcagaaaga	gcagagaaag	tccttctgtc	tcccagacc	tcccatgggc	tttgtgtctt	31260
tgctgaagac	aaccaactag	tgtggcaact	agtttgatat	cttttctctc	tataatagaa	31320
atgagatcat	ccaagttacg	ctggagaact	gggtgtgtgt	tctgtgtctc	tgtattcccc	31380
ctaccccttt	ctcctagaga	aattcattgt	agtttgggtta	gttggaagtc	ctggtaaggt	31440
ggcaggctct	gtggtatagt	gagggcctta	ttgctgggct	gtgggactcc	aaagtgtgct	31500
tcaagtcctc	cgaggactca	acctagatcc	caatctttcc	aaccaccctt	tctttccaac	31560
cctatgacat	taaattttca	gtgcttcaat	tttgagagaa	cattcagacc	atggtaaaat	31620
cacatgagca	ccaagttgga	ttcttagggg	aattctgaag	gaactgcttt	taaaaaataa	31680
cagttttctc	gatgaggttg	catttgaacc	agctgtttag	aactgaacct	cctcctcccc	31740
tgtgaatggc	ttgaacaggt	ggtttatttt	ccagaagcaa	catagagctg	gattaaattg	31800
ttattagggg	actaaattat	ctgttgacta	tactgctaac	ctcaacctca	aaatgtagta	31860
gtggctacca	gaccacccca	acacctctct	gattaaagcc	atcaagagga	cagttagagga	31920
cagaagaatg	cctatttccc	ttcaagacat	ttccagagag	tcctaatgt	ttctacttag	31980
ctgtgactgc	ctggaactta	gtacaggctt	catttagctg	caagggaagc	tgacaagtga	32040
aatcttccag	tttgacatcc	tcaataaaat	tgaaggtctg	agggatgaaa	ctattctgag	32100
aacagtgagg	gtgttcacac	catcttctga	atataaatgc	tgaagattac	taaatggata	32160
tgggtgaattt	taccacagtc	caaaccaaca	gagcaggtcg	gggatgtggc	tgagttggta	32220
cagtacactt	gcctagcata	gaacagctgc	tgtggtaaca	aacattgtta	caccacacac	32280
tcgggtgtgtg	gatgcaggag	aattagacat	ttcaaggcca	tcctcagctc	tgcatgtagt	32340
tcaaggccac	ctgggggttaa	ctgattttcca	ccttactttt	tgaggcagtg	tcactaaacc	32400
tgtagttcac	tgatgaagct	acagtggccc	acaaactcca	gcagccctcc	cagggctgtg	32460
actaccttta	aacaacaaca	aaacaaaaag	tgctggagat	ccaaactcaa	gttcttatcc	32520
tcaggggcaag	cacttttactg	actgggcttc	caggcctgct	ctattgtctg	ggcaagactc	32580
gctataacta	ccfacaattc	ctcttcaaca	tctatttggg	ttctcttttt	ctttgttttt	32640
gagacaaggt	ttctctgtgt	atctatccct	ggctgtgcta	gaactcacta	tgtaaaccag	32700
gctgtcctcg	aactcaaaga	ggtctgctcg	cttctgcctc	atgagtgtcg	ggattaaagg	32760
tgtaccacca	cacccctacc	tattgggttc	ttaaaggcca	ctttccaaaa	ctgatgaagg	32820
agagttaaaa	taaggagaat	ttttgtaaaa	ctaactgtta	atgtgaactg	tgaatgtatg	32880
ccccaaaatc	ctacatatct	acaaaataca	cgagtttttc	gttatttttag	tcacctctc	32940
tgggttgctt	ggttttactc	tttagtaatt	tttactaggc	aactttcagg	acagaatgac	33000
agttcttgaa	gtttaatctt	agttagaaaa	caggcaatag	tgagaaagct	gtgtcacagt	33060
gttgttctgc	aaccactcca	gatctattcc	ctcccaatgt	ctatggagtg	cattgtctgc	33120
ttaggtgctg	aataaggaca	tggcaacatt	tcttatctac	tgtaaaggcaa	aattggtgag	33180

cagtgtcact	aagcctgtag	ttcactgatg	aagctagagt	gcccactagt	gagccactat	33240
cactagttag	cacactagta	atagtgtgaa	agaaagttag	ttccactgtc	ccatgaacac	33300
accacacctc	atctctctta	ccttgacact	ggttcagatt	tcggcagatg	cagttagcat	33360
ggtgcttaag	aagtctgaag	gatagctctg	ggggatgggt	gcacatgaag	gtcacatctt	33420
tttttaatta	ggtgatttaa	atgggtttgt	gcacatgagt	gcaacaccca	tacaggccag	33480
aagggggcat	aagatccccg	gctctagagt	gtcaggctgt	tgtgaactgc	ctggaatagg	33540
tgttgggaac	tgatcttggg	tcccctggaa	gagcattgag	tacatgtcat	cactatctct	33600
ccagcctcac	actcttatcg	gtatgtctca	tttgtggggg	caatttgggg	agtccccact	33660
ctggatcagg	tactgtgaaa	tatggagttg	gcggggctgg	atccctgtct	tgccaccagc	33720
cagctgagga	aagctgtcaa	ttgtcttctt	gtgtctgcct	cagtttctta	gaaactagaa	33780
aggaaaaatg	gatggtatca	ctaagttcag	ccttccattg	taaggatcca	ttgaagtagt	33840
tgggtgtgat	tactcacgct	ggtgcccctc	cccttctgag	ctgcaagcat	cagctgtttg	33900
accagcagct	tctgtgctcc	gacaggaagc	agtgaggaaac	tgggctgcaa	gatgatccca	33960
gttctagcat	ttgctgcaac	cccccttctg	cacatctgtt	tccacatttt	atttcatcct	34020
tgagcacata	accttttcat	tttgatacat	gcttttcttt	acatagcctg	ggccagcctt	34080
gaactcatgc	tcctcctgcc	ataattcctt	aaggactgac	aatgacagtg	gacatgcacc	34140
gccaggctcc	actgctccta	gtacttttgt	tatgggtctt	ctatgtctgt	gatgtgtgga	34200
atgtgacact	gcacaccagg	cattgtggga	tgaagtagaa	catgttcatg	cacacaaaga	34260
tcaatcccaa	acagcatctg	cccacccttc	ccctctgccc	cttcccccg	tggtgttag	34320
cattctcttg	tgtagtgtcg	ggtcgcccc	ttctctgtct	ctaaacattg	aaacaagggg	34380
aacagaccca	tacataactc	caacacagcg	atctggtcca	agtctggatg	tagaacctt	34440
ggctctgtct	ttttctcttc	ttcccagagc	tgtcccagtc	gcctcccttt	ctaattggct	34500
ggtgctcatc	taacttgatg	tatatgttct	tttccctggc	tgttttatga	ctggcctgtc	34560
ccagtcatta	gtgcatctgg	tgtagaagc	taagctcaac	ttggcctcac	agttcttgatg	34620
ttcaggacac	atggggttat	ggctggatcc	tgtgtcaagc	acatcacttc	tttctgaacc	34680
cacacatctt	aagacatggg	cattgtcaca	tggtgacag	cagtacattc	ttgtatgtag	34740
ttttctctca	agtgttttgg	ttacatggcc	ctaaagccta	ggactgtctg	tcttcaatca	34800
tgtctgccac	ctgctgcccc	gcaagcccaa	gttgtggttc	ttcctgtcta	atctgcctct	34860
tttatcttta	gccctcctcc	ataagcctct	tttccactga	gctctggggt	attcagatta	34920
cagctcccta	ctgctttcct	gagaccctaa	gccgctccac	gatgcacagc	agcactttcc	34980
agcagtccta	tggagatgca	gtgtcgagat	gatatgagct	gtgctgtata	tgtaaaatgc	35040
atggtagaat	ctgaaaaacca	agcacaaaaa	aataaaaaaa	aagttcatta	atatttatag	35100
caatcatgta	atgttttgaa	aatacgtgcc	taataaaatt	attttccact	ttttaaaaat	35160
gtgggtacta	tgatgtttaa	tagcgcatat	tcatgtggca	tatttacctc	aatggggcat	35220
ggaaaacaag	acgacaaaag	gtgggatatt	tctctcaact	ttcctagctg	caacttgtga	35280
taatttgtga	gacacattaa	gcactcaaga	agggtttgat	tctgtagaca	aaacccgggg	35340
tcagacagtc	cacagcttca	gcatagtgtg	ttcactatat	ctaaagtagg	ttttacttct	35400
ttaagattaa	ttttgttctt	agacagtttc	ccacatgcag	cacaaagcag	ttatgtttta	35460
attttatagg	ctacaatgac	ggtatttcaa	aaagtaagat	gtgggagtta	caccttgaat	35520
gatcacccaa	ggatgttgg	ccatcagcac	tgaatggccc	agaaccaatc	tacagcaaat	35580
ctgagagagc	agagacatgc	tgctccactg	tctgagggcc	cttgcccatg	gtggttgcta	35640
ttgacatctt	gacaggatct	aaaatcacct	aggtgtgctc	tgggcacctc	tggacatctc	35700
tgggagtctc	tagactggat	tgcttgaggt	tgatttaaaa	accctaactg	ggcagcacca	35760
ttcattggga	tggggtcctg	gcgtggataa	agattggagg	atgagggagc	accaggttcc	35820
acctctctgc	ctcctgactg	gatgcagcat	gatcagctcc	ctctcctgct	gtgacacaat	35880
ctctgtaccc	tcaaaactgaa	caaaacaagc	tactcgaata	gttgctttcg	ttagtcaact	35940
tgtcacagct	ttgtcacagt	gatgagaaaa	ataacacaca	ccccccaaat	gaggattgga	36000
tcgaggccct	gggcatgctg	ggtaaaactct	tcaccactaa	gctatatctc	taaccttttt	36060
tactttgaat	agggctctaa	ttgcccagat	tggccttaaa	atgataattc	tcctgtcttc	36120
agcttcccaa	gcagctggaa	ttacaaatgc	gagcactagt	atgtatatgt	agtatataaa	36180
taatatattg	tacatatatt	cacatgtgtg	tgtacatgtg	tgtacacaaat	tacaaatggg	36240
agccccagaa	tgatgtatac	acatacatat	aagtaggtaa	ttaatatata	catatttaca	36300
catattatac	tagtgtgtat	atatagaata	tataatatgt	attgcacata	tattcacata	36360
tgtgtgtaca	tgtccatgtc	acattgtgat	tgctagttaa	ataccacttt	tcctgtctct	36420
ctaactccaa	tagcaattga	acctgtgaat	tatggtaaac	ttcaggggat	tgaaaagcct	36480
gacctccaca	gaatcagcta	acgttagctg	ctactatgat	tccaagcagt	agttactgga	36540
agtggtttgct	cctgtattct	cttgggaagga	agaactagat	tgtagtcat	cagttcacag	36600
tcgggttatgg	tccccctggt	ggcctcatag	accagaaaagt	agagttgcac	gggttgattg	36660
taaagtgacc	cacctccac	aggttcacat	gcttgaacac	tgggtcccca	gatgatggag	36720
ctcttggggg	aggttggtga	gacctctag	gaggttagct	tacctggagg	aaatgggctt	36780
ctaggggaca	gacattgaag	atgataggtt	ttccctgttt	ctgccctggg	aatatatcca	36840

```

aattcctcac tcgaatgcc cagccctgag atgagccttc cccttcacag tggactctac 36900
cctcaaacc tgagccagag tagatccttc gtcctttgct agatgtttgg tcacggcagt 36960
gaggtaaaagt actgagcaca gcatcattgt tcccatttta cggatgggaa aactgagcct 37020
tgaggagatgc ccaaggctgt cagccttgag tgtttaaaag ctgcaaggat tggatgcatt 37080
tgtcctatat cagggaaaca tgatggggat ggagtctggg tgctgaggca ctgagttggc 37140
aagaggggaag gcctggtgtt cactcaaagt cattaaggac caattgtgtc tgcaatcctg 37200
tgccctccct tgaacaatag gtaagggtca gtgtgagccc tgattactcc cagcagaagg 37260
atacatctgc tttggagacc aaagtccctt cactgtagaa actaggtcct tcaaggtctc 37320
agataaaaag acaatgggtt ttgtgctaatt ttccacccaa tgggtgtgtg ggtcacacta 37380
ctcactgggc caacatggtg gtctggacag tcacacagga tgaagtggg aaaggcaagt 37440
ctccccggca cctccctat cgtcactgca aggccaccct gactaagagt cccctcttca 37500
atgctggccc tacgaatatc ttatcactct tcgtgtttac aagatttctc tccttggaag 37560
gtgtgatgtg gacagtgaag ctctcaacaa cccctactca cctaagacct agacaagaga 37620
gtctgggatg gccatgtatg tactccttca gtaattgaca gccattttct ttgtctagga 37680
agtcttctta caagcttccg caccataacg gtatcggtcg tcctgattta cagacatgtc 37740
attgggactg atatctgcaa caaagagcag atctccagtc actcatctca ctccagcttg 37800
ttcacagAAC aaaaaggagt tgaagggagg tcttcatcat tgggtgttcc cttccatgga 37860
gcataggaca ggcattggtg ccagaacctg cccagcttct agctctccaa gcctcatgct 37920
ttcctgtctc naaaaaaaaa aaaaaaaaaa 37950

```

<210> 184

<211> 1381

<212> DNA

<213> Mus musculus

<400> 184

```

gagccgagag gatgctgctg tccctgggtg tccacacgta ctct atg cgc tac ctg 56
Met Arg Tyr Leu
1
ctc ccc agc gtc ctg ttg ctg ggc tgg gcg ccc acc tac ctg ctg gcc 104
Leu Pro Ser Val Leu Leu Leu Gly Ser Ala Pro Thr Tyr Leu Leu Ala
5 10 15 20
tgg acg ctg tgg cgg gtg ctc tcc gcg ctg atg ccc gcc cgc ctg tac 152
Trp Thr Leu Trp Arg Val Leu Ser Ala Leu Met Pro Ala Arg Leu Tyr
25 30 35
cag cgc gtg gac gac cgg ctt tac tgc gtc tac cag aac atg gtg ctc 200
Gln Arg Val Asp Arg Leu Tyr Cys Val Tyr Gln Asn Met Val Leu
40 45 50
ttc ttc ttc gag aac tac acc ggg gtc cag ata ttg cta tat gga gat 248
Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln Ile Leu Leu Tyr Gly Asp
55 60 65
ttg cca aaa aat aaa gaa aat gta ata tat cta gcg aat cat caa agc 296
Leu Pro Lys Asn Lys Glu Asn Val Ile Tyr Leu Ala Asn His Gln Ser
70 75 80
aca gtt gac tgg att gtt gcg gac atg ctg gct gcc aga cag gat gcc 344
Thr Val Asp Trp Ile Val Ala Asp Met Leu Ala Ala Arg Gln Asp Ala
85 90 95 100
cta gga cat gtg cgc tac gta ctg aaa gac aag tta aaa tgg ctt ccg 392
Leu Gly His Val Arg Tyr Val Leu Lys Asp Lys Leu Lys Trp Leu Pro
105 110 115
ctg tat ggg ttc tac ttt gct cag cat gga gga att tat gta aaa cga 440
Leu Tyr Gly Phe Tyr Phe Ala Gln His Gly Gly Ile Tyr Val Lys Arg
120 125 130
agt gcc aaa ttt aat gat aaa gaa atg aga agc aag ctg cag agc tat 488
Ser Ala Lys Phe Asn Asp Lys Glu Met Arg Ser Lys Leu Gln Ser Tyr
135 140 145
gtg aac gca gga aca ccg atg tat ctt gtg att ttc cca gag gga aca 536
Val Asn Ala Gly Thr Pro Met Tyr Leu Val Ile Phe Pro Glu Gly Thr
150 155 160
agg tat aat gca aca tac aca aaa ctc ctt tca gcc agt cag gca ttt 584
Arg Tyr Asn Ala Thr Tyr Thr Lys Leu Leu Ser Ala Ser Gln Ala Phe
165 170 175 180
gct gct cag cgg ggc ctt gca gta tta aaa cac gta ctg aca cca aga 632

```

Ala Ala Gln Arg Gly Leu Ala Val Leu Lys His Val Leu Thr Pro Arg
 185 190 195
 ata aag gcc act cac gtt gct ttt gat tct atg aag agt cat tta gat 680
 Ile Lys Ala Thr His Val Ala Phe Asp Ser Met Lys Ser His Leu Asp
 200 205 210
 gca att tat gat gtc aca gtg gtt tat gaa ggg aat gag aaa ggt tca 728
 Ala Ile Tyr Asp Val Thr Val Val Tyr Glu Gly Asn Glu Lys Gly Ser
 215 220 225
 gga aaa tac tca aat cca cca tcc atg act gag ttt ctc tgc aaa cag 776
 Gly Lys Tyr Ser Asn Pro Pro Ser Met Thr Glu Phe Leu Cys Lys Gln
 230 235 240
 tgc cca aaa ctt cat att cac ttt gat cgt ata gac aga aat gaa gtt 824
 Cys Pro Lys Leu His Ile His Phe Asp Arg Ile Asp Arg Asn Glu Val
 245 250 255 260
 cca gag gaa caa gaa cac atg aaa aag tgg ctt cat gag cgc ttt gag 872
 Pro Glu Glu Gln Glu His Met Lys Lys Trp Leu His Glu Arg Phe Glu
 265 270 275
 ata aaa gat agg ttg ctc ata gag ttc tat gat tca cca gat cca gaa 920
 Ile Lys Asp Arg Leu Leu Ile Glu Phe Tyr Asp Ser Pro Asp Pro Glu
 280 285 290
 aga aga aac aaa ttt cct ggg aaa agt gtt cat tcc aga cta agt gtg 968
 Arg Arg Asn Lys Phe Pro Gly Lys Ser Val His Ser Arg Leu Ser Val
 295 300 305
 aag aag act tta cct tca gtg ttg atc ttg ggg agt ttg act gcg gtc 1016
 Lys Lys Thr Leu Pro Ser Val Leu Ile Leu Gly Ser Leu Thr Ala Val
 310 315 320
 atg ctg atg acg gag tcc gga agg aaa ctg tac atg ggc acc tgg ttg 1064
 Met Leu Met Thr Glu Ser Gly Arg Lys Leu Tyr Met Gly Thr Trp Leu
 325 330 335 340
 tat gga acc ctc ctt ggc tgc ctg tgg ttt gtt att aaa gca taa 1109
 Tyr Gly Thr Leu Leu Gly Cys Leu Trp Phe Val Ile Lys Ala *
 345 350 355
 gcaagtagca ggctgcagtc acagtctctt attgatggct acacattgta tcacattggt 1169
 tcctgaatta aataaggagt tttcttggtt ttgttttttt tgttttggtt tgttctgttt 1229
 taagccttga tgattgaaca ctggataaag tcgagtcttg tgaccacagc caacatgcat 1289
 ttgatttggg gcaaacacat gtggcttttc aggtgctggg gttgctggag acatggaagc 1349
 taagtggagt ttatgctggt tttttttttt tt 1381
 <210> 185
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 4-14-107
 <221> allele
 <222> 24
 <223> polymorphic base A
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 4-14-107.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 4-14-107.mis2
 <400> 185
 ctaaacaacc accaaatgca tacagcaacc aggcaaatgc ctgatag
 <210> 186
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>

<221> allele
<222> 1..47
<223> polymorphic fragment 4-14-317
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-14-317.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-14-317.mis2
<400> 186
cataacatgc aaggtgggca agaaaaagag gtgggcacag ctcatga 47
<210> 187
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-14-35
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-14-35.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-14-35.mis2
<400> 187
atccaacaca gaaaccgcta aaaccaggca gaagctgtct gcagaga 47
<210> 188
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-20-149
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-20-149.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-20-149.mis2
<400> 188
tttttgctgt gtcttcaaag tgactcttgg tttattgcct gctaagg 47
<210> 189
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-20-77
<221> allele

```

<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-20-77.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-20-77.mis2
<400> 189
tgcaacatga agattctgaa gggactttgt tgtctgagaa cacatct      47
<210> 190
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-22-174
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-22-174.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-22-174.mis2
<400> 190
ggattgtgca gaagttgcct ttcattgtca aaaatgttaa tttgttt      47
<210> 191
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-22-176
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-22-176.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-22-176.mis2
<400> 191
attgtgcaga agttgccttt catattcaaa aatgttaatt tgtttgt      47
<210> 192
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-26-60
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23

```

<223> potential microsequencing oligo 4-26-60.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-26-60.mis2
<400> 192
gatgggaaag tgcattcttaa gacagtttagc aggccaagga gcgactt 47
<210> 193
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-26-72
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-26-72.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-26-72.mis2
<400> 193
catcttaaga cagtttagcag gccaaaggagc gactttaag ggtgagc 47
<210> 194
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-3-130
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-3-130.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-3-130.mis2
<400> 194
tattgggcct aaaacagtat tctataaagc ttaaattggt attact 47
<210> 195
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-38-63
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-38-63.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-38-63.mis2

<400> 195
tataagttat aagaaaatca ggcagaggct aaactttttt tttgttt 47
<210> 196
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-38-83
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-38-83.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-38-83.mis2
<400> 196
ggcagaggct aaactttttt tttgtttggc aatgctgttg agaatat 47
<210> 197
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-4-152
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-4-152.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-4-152.mis2
<400> 197
tactttcca ttgttcctga cttcgttatc ctatatataa acagaaa 47
<210> 198
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-4-187
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-4-187.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-4-187.mis2
<400> 198
tataaacaga aacatggatg agtaaaaaaa aaaaaaaaaa aaaaaaa 47
<210> 199
<211> 47

<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-4-288
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-4-288.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-4-288.mis2
<400> 199
ctgtcatcaa ctaattttca caagtaccta tgttttgatt tcatgta. 47
<210> 200
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-42-304
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-42-304.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-42-304.mis2
<400> 200
attatttaaa actatttatg taaccttatt ttcaggggtt tttaatt 47
<210> 201
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-42-401
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-42-401.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-42-401.mis2
<400> 201
taagaaagaa ttctgtgttc tggacaaagt ttaaaccac agagcca 47
<210> 202
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele

<222> 1..47
<223> polymorphic fragment 4-43-328
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-43-328.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-43-328.mis2
<400> 202
agaattctgt gttctggcca aagcttaaac ccacagagcc agtttaa 47
<210> 203
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-43-70
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-43-70.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-43-70.mis2
<400> 203
atcgctcca ttattctcaa aaagaccatg ggacacaaca caagaag 47
<210> 204
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-50-209
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-50-209.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-50-209.mis2
<400> 204
atatagagtg tgcattccctg acactgaaac tgaaggcttt atgggtt 47
<210> 205
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-50-293
<221> allele
<222> 24

<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-50-293.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-50-293.mis2
<400> 205
cctgagtcctc agggggctga caggggacag tttaaaacat tgatgaa 47
<210> 206
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-50-323
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-50-323.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-50-323.mis2
<400> 206
tttaaaacat tgatgaatct ttactactac aaaagggttc gatttag 47
<210> 207
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-50-329
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-50-329.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-50-329.mis2
<400> 207
acattgatga atctttatta ctacaaaagg gttcgattta ggctagc 47
<210> 208
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-50-330
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-50-330.mis1

<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-50-330.mis2
<400> 208
cattgatgaa tctttattac tacaaaaggg ttcgatttag gctagcc 47
<210> 209
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-52-163
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-52-163.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-52-163.mis2
<400> 209
gaacaggata ttcttaacta ccaaagaatt ttacacatct attgttt 47
<210> 210
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-52-88
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-52-88.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-52-88.mis2
<400> 210
tccatgtcat tattattcaa aagcttaaaa aatacacaag gtgaaaa 47
<210> 211
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-53-258
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-53-258.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-53-258.mis2
<400> 211

gagaaatcat gcagagagaa tgcattctca ctcaaatttt aacctaa 47
<210> 212
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-54-283
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-54-283.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-54-283.mis2
<400> 212
aagtagtttt tcacactttc tctatgatac aatcgatggc ttaatct 47
<210> 213
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-54-388
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-54-388.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-54-388.mis2
<400> 213
ctctctatcg tatacatctt tacacacgct gcagcgccaa gactcca 47
<210> 214
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-55-70
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-55-70.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-55-70.mis2
<400> 214
tattaagaac ctaggtttta aaaaactctc tatcgtatac atcttta 47
<210> 215
<211> 47
<212> DNA

<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-55-95
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-55-95.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-55-95.mis2
<400> 215
ctctctatcg tatacatctt tacacacgct gcagcgccaa gactcca 47
<210> 216
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-56-159
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-56-159.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-56-159.mis2
<400> 216
aagttttcct tctcttctgt agacgtctcc atgttacagt caactat 47
<210> 217
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-56-213
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-56-213.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-56-213.mis2
<400> 217
atggctcatg ttcactctgg ttcaccttca gaggagtttg atatttt 47
<210> 218
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47

<223> polymorphic fragment 4-58-289
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-58-289.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-58-289.mis2
<400> 218
catacctgca gcctgctttt ggtgaggggt gactacttta cctgcaa 47
<210> 219
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-58-318
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-58-318.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-58-318.mis2
<400> 219
tgactacttt acctgcaata tttatttgca agtttatttc ttccttt 47
<210> 220
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-60-266
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-60-266.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-60-266.mis2
<400> 220
aacaggacca agacactgca ttagataaag tttcagtatt tcttagc 47
<210> 221
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-60-293
<221> allele
<222> 24
<223> polymorphic base C

<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-60-293.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-60-293.mis2
<400> 221
aagtttcagt atttcttagc agacgaagcc agcaggaagt cctccta 47
<210> 222
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-84-241
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-84-241.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-84-241.mis2
<400> 222
gaaaaaaaa tagtgactgc cacggtgaat aattcagttc ttcagaa 47
<210> 223
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-84-262
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-84-262.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-84-262.mis2
<400> 223
acggtgaata attcagttct tcaaaagcag caacatgatc tcatgga 47
<210> 224
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-86-206
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-86-206.mis1
<221> primer_bind

<222> 25..47
<223> complement potential microsequencing oligo 4-86-206.mis2
<400> 224
gtattcaaat caggacacac cacaaatggc atctacacgt taacatt 47
<210> 225
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-86-309
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-86-309.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-86-309.mis2
<400> 225
tggctctagg caggccactt tagagagtga ggaaccagag agcagaa 47
<210> 226
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-88-349
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-88-349.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-88-349.mis2
<400> 226
gaaactaaaa gacaatattc agtgtgagat tttccaagtt ctttatg 47
<210> 227
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-89-87
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-89-87.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-89-87.mis2
<400> 227
ttcttcctg aacgctgggt tcacatagtt tttgtgtga gaataga 47

<210> 228
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-123-184
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-123-184.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-123-184.mis2
<400> 228
ccagcccaga acattcacca gctgggccaa gagttctgct gggttttt 47
<210> 229
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-128-202
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-128-202.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-128-202.mis2
<400> 229
aatgtctggtt tcttagagaa ctgaaacaca cacacatata tacacac 47
<210> 230
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-128-275
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-128-275.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-128-275.mis2
<400> 230
acaccctac ctacatgtg tagacaaatg tatgcatata tgtctct 47
<210> 231
<211> 47
<212> DNA
<213> Homo Sapiens

<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-128-313
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-128-313.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-128-313.mis2
<400> 231
tatgtctcta gacagatata cataagattc tatttggcat agaaaaa 47
<210> 232
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-128-60
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-128-60.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-128-60.mis2
<400> 232
gcactgtgac ccagggcgcta ggtccctctt acagtgcac tccgaca 47
<210> 233
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-12907-295
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-12907-295.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-12907-295.mis2
<400> 233
gctatatggc attatatctc cacagggcag acctgatgta caagatg 47
<210> 234
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-130-58

<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-130-58.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-130-58.mis2
<400> 234
aaagcaaaag agcttcaaaa atacttcagg agtgtgcata tggcgag 47
<210> 235
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-134-362
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-134-362.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-134-362.mis2
<400> 235
caaaacactc atgttagtta gatgattatt cctattacaa agataag 47
<210> 236
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-140-130
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-140-130.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-140-130.mis2
<400> 236
tggtcaaaaag cagctacaga ccacatgtaa acaattgagc atggctg 47
<210> 237
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1462-238
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind

<222> 1..23
<223> potential microsequencing oligo 99-1462-238.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1462-238.mis2
<400> 237
ccctttcaag gttagtaact catgtgctgt gtttctgctt cagaagg 47
<210> 238
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-147-181
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-147-181.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-147-181.mis2
<400> 238
gtgtcatgaa aaagagcatg ataaaaagaa aaacttaaat ctttata 47
<210> 239
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1474-156
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1474-156.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1474-156.mis2
<400> 239
cttgactca taagttaaat attgataaca agaagaaata tggactt 47
<210> 240
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1474-359
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1474-359.mis1
<221> primer_bind
<222> 25..47

<223> complement potential microsequencing oligo 99-1474-359.mis2
<400> 240
aaaaaaaatc aaattattgt accaaattcc ctaatatcag atgtgta 47
<210> 241
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1479-158
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1479-158.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1479-158.mis2
<400> 241
tttaaaaatc cacttgtaat cgccgctaatt tggagtgtat attcagg 47
<210> 242
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1479-379
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1479-379.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1479-379.mis2
<400> 242
gtagagctgt gtactgaggt cagagaagca gctcatggta cagcctt 47
<210> 243
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-148-129
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-148-129.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-148-129.mis2
<400> 243
ttcatatcta tacaaataat tttaaattta atacataggg ctgcaaa 47
<210> 244

<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-148-132
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-148-132.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-148-132.mis2
<400> 244
atatctatac aaataatttt gaacttaata catagggctg caaaaca 47
<210> 245
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-148-139
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-148-139.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-148-139.mis2
<400> 245
tacaaataat tttgaattta atacataggg ctgcaaaaca aggttga 47
<210> 246
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-148-140
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-148-140.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-148-140.mis2
<400> 246
acaaataatt ttgaatttaa tacatagggc tgcaaaaca ggttgat 47
<210> 247
<211> 47
<212> DNA
<213> Homo Sapiens
<220>

<221> allele
<222> 1..47
<223> polymorphic fragment 99-148-182
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-148-182.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-148-182.mis2
<400> 247
ttgatgttga tatgggcaac tgtatgttgg atggtcccaa agcattc 47
<210> 248
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-148-366
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-148-366.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-148-366.mis2
<400> 248
tccttgtaa aggtctctcc ctggtgctca cggctgccgc ctcaaag 47
<210> 249
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-148-76
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-148-76.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-148-76.mis2
<400> 249
tgatagaatg ccttctgaa ttactactct tgatggcttc ataaaac 47
<210> 250
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1480-290
<221> allele

<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1480-290.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1480-290.mis2
<400> 250
tgcaccatct tcaccacaac cccgggcaac cactgatcct tttactg 47
<210> 251
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1481-285
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1481-285.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1481-285.mis2
<400> 251
tcccataacc tgttttgctt ctcgctctaa cctcaagatg gtataaa 47
<210> 252
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1484-101
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1484-101.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1484-101.mis2
<400> 252
aaaaagatca aatataagca tgtaactcct ctccttaaaa tctcagt 47
<210> 253
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1484-328
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23

<223> potential microsequencing oligo 99-1484-328.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1484-328.mis2
<400> 253
ggacacgtgg tcatgaggag ttggaaggga ttcagttttc agatccc 47
<210> 254
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1485-251
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1485-251.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1485-251.mis2
<400> 254
gattgccttg atatatgctc ccagagaacc aagaatgtcc ccttttc 47
<210> 255
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1490-381
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1490-381.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1490-381.mis2
<400> 255
tgcacagtgg aaataccatg tcacggtacg ctactgtgca tctcttc 47
<210> 256
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1493-280
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1493-280.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1493-280.mis2

<400> 256
ggatgacaga gttatgttgg aggaatgggg ttggctgct tgttttt 47
<210> 257
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-151-94
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-151-94.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-151-94.mis2
<400> 257
attgagatca ttgataagga aatattctaa aatttcaaaa tctatat 47
<210> 258
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-211-291
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-211-291.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-211-291.mis2
<400> 258
ctggttatat cagactgacc ttcattgttt caacagggtca atgcctt 47
<210> 259
<211> 45
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..45
<223> polymorphic fragment 99-213-37
<221> allele
<222> 23
<223> polymorphic base T
<221> primer_bind
<222> 1..22
<223> potential microsequencing oligo 99-213-37.mis1
<221> primer_bind
<222> 24..45
<223> complement potential microsequencing oligo 99-213-37.mis2
<400> 259
gtgcttccgg ctgcaggact gttggaggac tccagtgtct gacag 45
<210> 260
<211> 47

<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-221-442
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-221-442.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-221-442.mis2
<400> 260
tgcccttgta gatatgcatg ggaattccat gacctagcca gacgaat 47
<210> 261
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-222-109
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-222-109.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-222-109.mis2
<400> 261
caggtgagga gtgctggatt ggccacgata tgaatttctt cagcagt 47
<210> 262
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-14-107, variant version of SEQ ID185
<221> allele
<222> 24
<223> base G ; A in SEQ ID185
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-14-107.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-14-107.mis2
<400> 262
ctaaacaacc accaaatgca tacggcaacc aggcaaatgc ctgatag 47
<210> 263
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele

<222> 1..47
<223> polymorphic fragment 4-14-317, variant version of SEQ ID186
<221> allele
<222> 24
<223> base G ; A in SEQ ID186
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-14-317.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-14-317.mis2
<400> 263
cataacatgc aaggtgggca agagaaagag gtgggcacag ctcatga 47
<210> 264
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-14-35, variant version of SEQ ID187
<221> allele
<222> 24
<223> base T ; C in SEQ ID187
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-14-35.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-14-35.mis2
<400> 264
atccaacaca gaaaccgcta aaatcaggca gaagctgtct gcagaga 47
<210> 265
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-20-149, variant version of SEQ ID188
<221> allele
<222> 24
<223> base T ; C in SEQ ID188
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-20-149.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-20-149.mis2
<400> 265
tttttgctgt gtcttcaaag tgattcttgg tttattgcct gctaagg 47
<210> 266
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-20-77, variant version of SEQ ID189
<221> allele
<222> 24

<223> base T ; A in SEQ ID189
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-20-77.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-20-77.mis2
<400> 266
tgcaacatga agattctgaa gggctcttgt tgtctgagaa cacatct 47
<210> 267
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-22-174, variant version of SEQ ID190
<221> allele
<222> 24
<223> base C ; A in SEQ ID190
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-22-174.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-22-174.mis2
<400> 267
ggattgtgca gaagttgcct ttctgttca aaaatgttaa tttgttt 47
<210> 268
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-22-176, variant version of SEQ ID191
<221> allele
<222> 24
<223> base G ; A in SEQ ID191
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-22-176.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-22-176.mis2
<400> 268
attgtgcaga agttgccttt catgttcaaa aatgttaatt tgtttgt 47
<210> 269
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-26-60, variant version of SEQ ID192
<221> allele
<222> 24
<223> base G ; A in SEQ ID192
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-26-60.mis1

<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-26-60.mis2
<400> 269
gatgggaaag tgcattcttaa gacggtagc aggccaagga gcgactt 47
<210> 270
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-26-72, variant version of SEQ ID193
<221> allele
<222> 24
<223> base G ; A in SEQ ID193
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-26-72.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-26-72.mis2
<400> 270
catcttaaga cagtttagcag gccgaggagc gactttaag ggtgagc 47
<210> 271
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-3-130, variant version of SEQ ID194
<221> allele
<222> 24
<223> base G ; A in SEQ ID194
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-3-130.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-3-130.mis2
<400> 271
tattgggcct aaaacagtat tctgtaaagc ttaaattggt attaact 47
<210> 272
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-38-63, variant version of SEQ ID195
<221> allele
<222> 24
<223> base G ; A in SEQ ID195
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-38-63.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-38-63.mis2
<400> 272

tataagttat aagaaaatca ggccgaggct aaactttttt tttgttt 47
<210> 273
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-38-83, variant version of SEQ ID196
<221> allele
<222> 24
<223> base T ; G in SEQ ID196
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-38-83.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-38-83.mis2
<400> 273
ggcagaggct aaactttttt ttttttggc aatgctgttg agaatat 47
<210> 274
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-4-152, variant version of SEQ ID197
<221> allele
<222> 24
<223> base T ; C in SEQ ID197
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-4-152.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-4-152.mis2
<400> 274
tactttccca ttgttctga cttgttatc ctatatataa acagaaa 47
<210> 275
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-4-187, variant version of SEQ ID198
<221> allele
<222> 24
<223> base T ; A in SEQ ID198
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-4-187.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-4-187.mis2
<400> 275
tataaacaga aacatggatg agttaaaaaa aaaaaaaaaa aaaaaaa 47
<210> 276
<211> 47
<212> DNA

```

<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-4-288, variant version of SEQ ID199
<221> allele
<222> 24
<223> base C ; G in SEQ ID199
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-4-288.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-4-288.mis2
<400> 276
ctgtcatcaa ctaattttca caactaccta tgttttgatt tcattga      47
<210> 277
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-42-304, variant version of SEQ ID200
<221> allele
<222> 24
<223> base T ; C in SEQ ID200
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-42-304.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-42-304.mis2
<400> 277
attatttaaa actatttatg taatcttatt ttcaggggtt ttaatt      47
<210> 278
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-42-401, variant version of SEQ ID201
<221> allele
<222> 24
<223> base C ; A in SEQ ID201
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-42-401.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-42-401.mis2
<400> 278
taagaaagaa ttctgtgttc tggccaaagt ttaaaccac agagcca      47
<210> 279
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47

```


<223> polymorphic fragment 4-43-328, variant version of SEQ ID202
<221> allele -
<222> 24
<223> base T ; C in SEQ ID202
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-43-328.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-43-328.mis2
<400> 279
agaattctgt gttctggcca aagtttaaac ccacagagcc agttaa 47
<210> 280
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-43-70, variant version of SEQ ID203
<221> allele
<222> 24
<223> base C ; G in SEQ ID203
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-43-70.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-43-70.mis2
<400> 280
atcgccctca ttattctcaa aaacaccatg ggacacaaca caagaag 47
<210> 281
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-50-209, variant version of SEQ ID204
<221> allele
<222> 24
<223> base T ; C in SEQ ID204
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-50-209.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-50-209.mis2
<400> 281
atatagagtg tgcacccctg acattgaaac tgaaggcttt atggttt 47
<210> 282
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-50-293, variant version of SEQ ID205
<221> allele
<222> 24
<223> base T ; G in SEQ ID205

<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-50-293.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-50-293.mis2
<400> 282
cctgagtcctc agggggctga cagtggacag tttaaaacat tgatgaa 47
<210> 283
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-50-323, variant version of SEQ ID206
<221> allele
<222> 24
<223> base T ; C in SEQ ID206
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-50-323.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-50-323.mis2
<400> 283
tttaaaacat tgatgaatct ttattactac aaaagggttc gatttag 47
<210> 284
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-50-329, variant version of SEQ ID207
<221> allele
<222> 24
<223> base T ; C in SEQ ID207
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-50-329.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-50-329.mis2
<400> 284
acattgatga atctttatta ctataaaagg gttcgattta ggctagc 47
<210> 285
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-50-330, variant version of SEQ ID208
<221> allele
<222> 24
<223> base T ; A in SEQ ID208
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-50-330.mis1
<221> primer_bind

<222> 25..47
<223> complement potential microsequencing oligo 4-50-330.mis2
<400> 285
cattgatgaa tctttattac tactaaaggg ttcgatttag gctagcc 47
<210> 286
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-52-163, variant version of SEQ ID209
<221> allele
<222> 24
<223> base C ; A in SEQ ID209
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-52-163.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-52-163.mis2
<400> 286
gaacaggata ttcttaacta ccacagaatt ttacacatct attgttt 47
<210> 287
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-52-88, variant version of SEQ ID210
<221> allele
<222> 24
<223> base T ; C in SEQ ID210
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-52-88.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-52-88.mis2
<400> 287
tccatgtcat tattattcaa aagtttaaaa aatacacaag gtgaaaa 47
<210> 288
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-53-258, variant version of SEQ ID211
<221> allele
<222> 24
<223> base G ; A in SEQ ID211
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-53-258.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-53-258.mis2
<400> 288
gagaaatcat gcagagagaa tgcgttctca ctcaaatttt aacctaa 47

<210> 289
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-54-283, variant version of SEQ ID212
<221> allele
<222> 24
<223> base T ; A in SEQ ID212
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-54-283.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-54-283.mis2
<400> 289
aagtagtttt tcacactttc tctttgatac aatcgatggc ttaatct 47
<210> 290
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-54-388, variant version of SEQ ID213
<221> allele
<222> 24
<223> base C ; A in SEQ ID213
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-54-388.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-54-388.mis2
<400> 290
ctctctatcg tatacatctt tacccacgct gcagcgccaa gactcca 47
<210> 291
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-55-70, variant version of SEQ ID214
<221> allele
<222> 24
<223> base T ; A in SEQ ID214
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-55-70.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-55-70.mis2
<400> 291
tattaagaac ctaggtttta aaatactctc tatcgtatac atcttta 47
<210> 292
<211> 47
<212> DNA
<213> Homo Sapiens

<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-55-95, variant version of SEQ ID215
<221> allele
<222> 24
<223> base C ; A in SEQ ID215
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-55-95.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-55-95.mis2
<400> 292
ctctctatcg tatacatctt taccacgct gcagcgcaa gactcca 47
<210> 293
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-56-159, variant version of SEQ ID216
<221> allele
<222> 24
<223> base T ; C in SEQ ID216
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-56-159.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-56-159.mis2
<400> 293
aagttttcct tctcttctgt agatgtctcc atgttacagt caactat 47
<210> 294
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-56-213, variant version of SEQ ID217
<221> allele
<222> 24
<223> base G ; A in SEQ ID217
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-56-213.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-56-213.mis2
<400> 294
atggctcatg ttcactctgg ttgccttca gaggagtttg atatttt 47
<210> 295
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-58-289, variant version of SEQ ID218

<221> allele
<222> 24
<223> base C ; G in SEQ ID218
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-58-289.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-58-289.mis2
<400> 295
catacctgca gcctgctttt ggtcaggggt gactacttta cctgcaa 47
<210> 296
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-58-318, variant version of SEQ ID219
<221> allele
<222> 24
<223> base C ; A in SEQ ID219
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-58-318.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-58-318.mis2
<400> 296
tgactacttt acctgcaata tttctttgca agtttatttc ttccttt 47
<210> 297
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-60-266, variant version of SEQ ID220
<221> allele
<222> 24
<223> base T ; G in SEQ ID220
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-60-266.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-60-266.mis2
<400> 297
aacaggacca agacactgca ttatataaag tttcagtatt tcttagc 47
<210> 298
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-60-293, variant version of SEQ ID221
<221> allele
<222> 24
<223> base T ; C in SEQ ID221
<221> primer_bind

<222> 1..23
<223> potential microsequencing oligo 4-60-293.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-60-293.mis2
<400> 298
aagtttcagt atttcttagc agatgaagcc agcaggaagt cctccta 47
<210> 299
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-84-241, variant version of SEQ ID222
<221> allele
<222> 24
<223> base T ; G in SEQ ID222
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-84-241.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-84-241.mis2
<400> 299
gaaaaaaaaa tagtgactgc cactgtgaat aattcagttc ttcagaa 47
<210> 300
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-84-262, variant version of SEQ ID223
<221> allele
<222> 24
<223> base G ; A in SEQ ID223
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-84-262.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-84-262.mis2
<400> 300
acgggtgaata attcagttct tcagaagcag caacatgatc tcatgga 47
<210> 301
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-86-206, variant version of SEQ ID224
<221> allele
<222> 24
<223> base G ; A in SEQ ID224
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-86-206.mis1
<221> primer_bind
<222> 25..47

<223> complement potential microsequencing oligo 4-86-206.mis2
<400> 301
gtattcaaat caggacacac cacgaatggc atctacacgt taacatt 47
<210> 302
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-86-309, variant version of SEQ ID225
<221> allele
<222> 24
<223> base T ; A in SEQ ID225
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-86-309.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-86-309.mis2
<400> 302
tggctctagg caggccactt tagtgagtga ggaaccagag agcagaa 47
<210> 303
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-88-349, variant version of SEQ ID226
<221> allele
<222> 24
<223> base C ; G in SEQ ID226
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-88-349.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-88-349.mis2
<400> 303
gaaactaaaa gacaatatc agtctgagat ttccaagtt ctttatg 47
<210> 304
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-89-87, variant version of SEQ ID227
<221> allele
<222> 24
<223> base T ; C in SEQ ID227
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-89-87.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-89-87.mis2
<400> 304
ttcttcctg aacgctggtt tcatatagtt tttgtgtga gaataga 47
<210> 305

<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-123-184, variant version of SEQ ID228
<221> allele
<222> 24
<223> base C ; G in SEQ ID228
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-123-184.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-123-184.mis2
<400> 305
ccagcccaga acattcacca gctcggccaa gagttctgct gggtttt 47
<210> 306
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-128-202, variant version of SEQ ID229
<221> allele
<222> 24
<223> base C ; A in SEQ ID229
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-128-202.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-128-202.mis2
<400> 306
aatgtctgtt tcttagagaa ctgcaacaca cacacatata tacacac 47
<210> 307
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-128-275, variant version of SEQ ID230
<221> allele
<222> 24
<223> base G ; A in SEQ ID230
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-128-275.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-128-275.mis2
<400> 307
acaccctac ctcacatgtg taggcaaag tatgcatata tgtctct 47
<210> 308
<211> 47
<212> DNA
<213> Homo Sapiens
<220>

<221> allele
<222> 1..47
<223> polymorphic fragment 99-128-313, variant version of SEQ ID231
<221> allele
<222> 24
<223> base G ; A in SEQ ID231
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-128-313.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-128-313.mis2
<400> 308
tatgtctcta gacagatata catgagattc tatttggcat agaaaaa 47
<210> 309
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-128-60, variant version of SEQ ID232
<221> allele
<222> 24
<223> base T ; C in SEQ ID232
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-128-60.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-128-60.mis2
<400> 309
gcactgtgac ccaggcgcta ggttcctctt acagtgcac tccgaca 47
<210> 310
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-12907-295, variant version of SEQ ID233
<221> allele
<222> 24
<223> base G ; A in SEQ ID233
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-12907-295.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-12907-295.mis2
<400> 310
gctatatggc attatatctc cacggggcag acctgatgta caagatg 47
<210> 311
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-130-58, variant version of SEQ ID234
<221> allele

```

<222> 24
<223> base T ; C in SEQ ID234
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-130-58.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-130-58.mis2
<400> 311
aaagcaaaag agcttcaaaa atatttcagg agtgtgcata tggcgag 47
<210> 312
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-134-362, variant version of SEQ ID235
<221> allele
<222> 24
<223> base T ; G in SEQ ID235
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-134-362.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-134-362.mis2
<400> 312
caaaacactc atgttagtta gattattatt cctattacaa agataag 47
<210> 313
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-140-130, variant version of SEQ ID236
<221> allele
<222> 24
<223> base T ; C in SEQ ID236
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-140-130.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-140-130.mis2
<400> 313
tgttcaaaag cagctacaga ccatatgtaa acaattgagc atggctg 47
<210> 314
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1462-238, variant version of SEQ ID237
<221> allele
<222> 24
<223> base C ; G in SEQ ID237
<221> primer_bind
<222> 1..23

```

<223> potential microsequencing oligo 99-1462-238.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1462-238.mis2
<400> 314
ccctttcaag gttagtaact catctgctgt gtttctgctt cagaagg 47
<210> 315
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-147-181, variant version of SEQ ID238
<221> allele
<222> 24
<223> base G ; A in SEQ ID238
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-147-181.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-147-181.mis2
<400> 315
gtgtcatgaa aaagagcatg atagaaagaa aaacttaaatt ctttata 47
<210> 316
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1474-156, variant version of SEQ ID239
<221> allele
<222> 24
<223> base T ; G in SEQ ID239
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1474-156.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1474-156.mis2
<400> 316
cttgactca taagttaaatt atttataaca agaagaaata tggactt 47
<210> 317
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1474-359, variant version of SEQ ID240
<221> allele
<222> 24
<223> base G ; A in SEQ ID240
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1474-359.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1474-359.mis2

```

<400> 317
aaaaaaaatc aaattattgt accgaattcc ctaatatcag atgtgta
<210> 318
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1479-158, variant version of SEQ ID241
<221> allele
<222> 24
<223> base T ; C in SEQ ID241
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1479-158.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1479-158.mis2
<400> 318
tttaaaaatc cacttgtaat cgctgctaatt tggagtgtat attcagg
<210> 319
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1479-379, variant version of SEQ ID242
<221> allele
<222> 24
<223> base G ; A in SEQ ID242
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1479-379.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1479-379.mis2
<400> 319
gtagagctgt gtactgaggt caggaagca gctcatggta cagcctt
<210> 320
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-148-129, variant version of SEQ ID243
<221> allele
<222> 24
<223> base G ; A in SEQ ID243
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-148-129.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-148-129.mis2
<400> 320
ttcatatcta tacaaataat ttgtaattta atacataggg ctgcaaa
<210> 321
<211> 47

```

<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-148-132, variant version of SEQ ID244
<221> allele
<222> 24
<223> base T ; C in SEQ ID244
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-148-132.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-148-132.mis2
<400> 321
atatctatac aaataatttt gaatttaata catagggctg caaaaca 47
<210> 322
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-148-139, variant version of SEQ ID245
<221> allele
<222> 24
<223> base T ; C in SEQ ID245
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-148-139.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-148-139.mis2
<400> 322
tacaaataat ttgaattta atatataggg ctgcaaaaca aggttga 47
<210> 323
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-148-140, variant version of SEQ ID246
<221> allele
<222> 24
<223> base G ; A in SEQ ID246
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-148-140.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-148-140.mis2
<400> 323
acaaataatt ttgaatttaa tacgtagggc tgcaaaacaa ggttgat 47
<210> 324
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele

<222> 1..47
<223> polymorphic fragment 99-148-182, variant version of SEQ ID247
<221> allele
<222> 24
<223> base G ; A in SEQ ID247
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-148-182.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-148-182.mis2
<400> 324
ttgatgttga tatgggcaac tgtgtgttgg atgggcccaa agcattc 47
<210> 325
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-148-366, variant version of SEQ ID248
<221> allele
<222> 24
<223> base T ; G in SEQ ID248
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-148-366.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-148-366.mis2
<400> 325
tccttgtcaa aggtctctcc ctgttgctca cggctgccgc ctcaaag 47
<210> 326
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-148-76, variant version of SEQ ID249
<221> allele
<222> 24
<223> base T ; C in SEQ ID249
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-148-76.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-148-76.mis2
<400> 326
tgatagaatg ccttcttgaa ttattactct tgatggcttc ataaaac 47
<210> 327
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1480-290, variant version of SEQ ID250
<221> allele
<222> 24

```

<223> base T ; G in SEQ ID250
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1480-290.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1480-290.mis2
<400> 327
tgccaccatct tcaccacaac ccctggcaac cactgatcct ttactg      47
<210> 328
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1481-285, variant version of SEQ ID251
<221> allele
<222> 24
<223> base T ; G in SEQ ID251
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1481-285.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1481-285.mis2
<400> 328
tcccataacc tgttttgctt ctctctctaa cctcaagatg gtataaa      47
<210> 329
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1484-101, variant version of SEQ ID252
<221> allele
<222> 24
<223> base C ; A in SEQ ID252
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1484-101.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1484-101.mis2
<400> 329
aaaaagatca aatataagca tgtcactcct ctccttaaaa tctcagt      47
<210> 330
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1484-328, variant version of SEQ ID253
<221> allele
<222> 24
<223> base C ; G in SEQ ID253
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1484-328.mis1

```


<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1484-328.mis2
<400> 330
ggacacgtgg tcacgaggag tttcaaggga ttcagttttc agatccc 47
<210> 331
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1485-251, variant version of SEQ ID254
<221> allele
<222> 24
<223> base T ; G in SEQ ID254
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1485-251.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1485-251.mis2
<400> 331
gattgccttg atatatgctc ccatagaacc aagaatgtcc ctttttc 47
<210> 332
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1490-381, variant version of SEQ ID255
<221> allele
<222> 24
<223> base T ; C in SEQ ID255
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1490-381.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1490-381.mis2
<400> 332
tgcacagtgg aaataccatg tcacggtacg ctactgtgca tctcttc 47
<210> 333
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1493-280, variant version of SEQ ID256
<221> allele
<222> 24
<223> base G ; A in SEQ ID256
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1493-280.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1493-280.mis2
<400> 333

ggatgacaga gtattgttgg agggatgggg ttggctgct tgttttt 47
<210> 334
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-151-94, variant version of SEQ ID257
<221> allele
<222> 24
<223> base G ; A in SEQ ID257
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-151-94.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-151-94.mis2
<400> 334
attgagatca ttgataagga aatgttctaa aatttcaaaa tctatat 47
<210> 335
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-211-291, variant version of SEQ ID258
<221> allele
<222> 24
<223> base G ; A in SEQ ID258
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-211-291.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-211-291.mis2
<400> 335
ctgggttatat cagactgacc ttcgtgtttt caacagggtca atgcctt 47
<210> 336
<211> 46
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..46
<223> polymorphic fragment 99-213-37, variant version of SEQ ID259
<221> allele
<222> 23
<223> base GC ; T in SEQ ID259
<221> primer_bind
<222> 1..22
<223> potential microsequencing oligo 99-213-37.mis1
<221> primer_bind
<222> 24..46
<223> complement potential microsequencing oligo 99-213-37.mis2
<400> 336
gtgcttcggg ctgcaggact gtgcggagga ctccagtgtc tgacag 46
<210> 337
<211> 47
<212> DNA

```

<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-221-442, variant version of SEQ ID260
<221> allele
<222> 24
<223> base C ; A in SEQ ID260
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-221-442.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-221-442.mis2
<400> 337
tgcctttgta gatatgcatg ggacttccat gacctagcca gacgaat      47
<210> 338
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-222-109, variant version of SEQ ID261
<221> allele
<222> 24
<223> base T ; C in SEQ ID261
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-222-109.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-222-109.mis2
<400> 338
caggtgagga gtgctggatt ggctacgata tgaatttctt cagcagt      47
<210> 339
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 185, SEQ 262, SEQ 186, SEQ 263,
SEQ 187, SEQ 264
<400> 339
tctaacctct catccaac      18
<210> 340
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> upstream amplification primer for SEQ 188, SEQ 265, SEQ 189, SEQ 266
<400> 340
gttatcgtga gactttttc      19
<210> 341
<211> 18
<212> DNA
<213> Homo Sapiens
<220>

```

<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 190, SEQ 267, SEQ 191, SEQ 268
<400> 341
tgctggtgct gtgataac 18
<210> 342
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 192, SEQ 269, SEQ 193, SEQ 270
<400> 342
tacagccctg taagacac 18
<210> 343
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> upstream amplification primer for SEQ 194, SEQ 271
<400> 343
cagtatgttc aatgcacag 19
<210> 344
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 195, SEQ 272, SEQ 196, SEQ 273
<400> 344
aaaacatcga catgggac 18
<210> 345
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 197, SEQ 274, SEQ 198, SEQ 275,
SEQ 199, SEQ 276
<400> 345
agcatttcga gtcattgtg 18
<210> 346
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 200, SEQ 277, SEQ 201, SEQ 278
<400> 346
ccctctttcc tcatgtag 18
<210> 347
<211> 19
<212> DNA
<213> Homo Sapiens
<220>

<221> primer_bind
<222> 1..19
<223> upstream amplification primer for SEQ 202, SEQ 279, SEQ 203, SEQ 280
<400> 347
taactcgtaa acagagaac 19
<210> 348
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 204, SEQ 281, SEQ 205, SEQ 282,
SEQ 206, SEQ 283, SEQ 207, SEQ 284, SEQ 208, SEQ 285
<400> 348
gcgtattgaa gctctttg 18
<210> 349
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 209, SEQ 286, SEQ 210, SEQ 287
<400> 349
aacacgggga ttttaggc 18
<210> 350
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> upstream amplification primer for SEQ 211, SEQ 288
<400> 350
cacatactaa ggctaattg 19
<210> 351
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 212, SEQ 289, SEQ 213, SEQ 290
<400> 351
gttgctggaa cctatttg 18
<210> 352
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 214, SEQ 291, SEQ 215, SEQ 292
<400> 352
tcgatggctt aatctacc 18
<210> 353
<211> 18
<212> DNA
<213> Homo Sapiens
<220>

```
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 216, SEQ 293, SEQ 217, SEQ 294
<400> 353
aaagaggagt aaatgggg                                     18
<210> 354
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 218, SEQ 295, SEQ 219, SEQ 296
<400> 354
tccccacagc taagagcc                                     18
<210> 355
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 220, SEQ 297, SEQ 221, SEQ 298
<400> 355
atacctaatt tcaggggg                                     18
<210> 356
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> upstream amplification primer for SEQ 222, SEQ 299, SEQ 223, SEQ 300
<400> 356
ttaacagagt accttgagg                                     19
<210> 357
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 224, SEQ 301, SEQ 225, SEQ 302
<400> 357
gtacagcctt ttgcttac                                     18
<210> 358
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 226, SEQ 303
<400> 358
aacgtgtcat agaaagcc                                     18
<210> 359
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
```

<222> 1..19
<223> upstream amplification primer for SEQ 227, SEQ 304
<400> 359
gctgatgagt tagataacc 19
<210> 360
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 228, SEQ 305
<400> 360
aaagccagga ctagaagg 18
<210> 361
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 229, SEQ 306, SEQ 230, SEQ 307,
SEQ 231, SEQ 308, SEQ 232, SEQ 309
<400> 361
gaccaggggt taagttag 18
<210> 362
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 233, SEQ 310
<400> 362
tctgttagga cctgtgag 18
<210> 363
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> upstream amplification primer for SEQ 234, SEQ 311
<400> 363
ccataacagc tagtacaac 19
<210> 364
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 235, SEQ 312
<400> 364
tggaaggtta ctcagaag 18
<210> 365
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind

<222> 1..19
<223> upstream amplification primer for SEQ 236, SEQ 313
<400> 365
agagcatagt ataaagcag 19
<210> 366
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> upstream amplification primer for SEQ 237, SEQ 314
<400> 366
ctagaagtag ctttaacag 19
<210> 367
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> upstream amplification primer for SEQ 238, SEQ 315
<400> 367
gcagccaatc ttatatattc 19
<210> 368
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> upstream amplification primer for SEQ 239, SEQ 316, SEQ 240, SEQ 317
<400> 368
aagggtgtag agtagaaag 19
<210> 369
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 241, SEQ 318, SEQ 242, SEQ 319
<400> 369
caactgacac tataaccc 18
<210> 370
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 243, SEQ 320, SEQ 244, SEQ 321,
SEQ 245, SEQ 322, SEQ 246, SEQ 323, SEQ 247, SEQ 324, SEQ 248, SEQ 325, SEQ
249, SEQ 326
<400> 370
cagtggagtg tttatgtg 18
<210> 371
<211> 19
<212> DNA
<213> Homo Sapiens
<220>

<221> primer_bind
<222> 1..19
<223> upstream amplification primer for SEQ 250, SEQ 327
<400> 371
ttgcacaaaa ggtatagag 19
<210> 372
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> upstream amplification primer for SEQ 251, SEQ 328
<400> 372
aggctccct tttgagttg 19
<210> 373
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 252, SEQ 329, SEQ 253, SEQ 330
<400> 373
atcctttcta gctgggag 18
<210> 374
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> upstream amplification primer for SEQ 254, SEQ 331
<400> 374
gtttaagaat gtgtgatggg 20
<210> 375
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> upstream amplification primer for SEQ 255, SEQ 332
<400> 375
aaggcaacag cgttgtgac 19
<210> 376
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 256, SEQ 333
<400> 376
ttttgggggt tttcagtg 18
<210> 377
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind

<222> 1..18
<223> upstream amplification primer for SEQ 257, SEQ 334
<400> 377
aacacaacag caaatccc 18
<210> 378
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 258, SEQ 335
<400> 378
tccttacttg taaccccc 18
<210> 379
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> upstream amplification primer for SEQ 259, SEQ 336
<400> 379
atactggcag cgtgtgcttc 20
<210> 380
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> upstream amplification primer for SEQ 260, SEQ 337
<400> 380
ccctttttct tcactgttc 19
<210> 381
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> upstream amplification primer for SEQ 261, SEQ 338
<400> 381
aggggagatg agggaagtgtg 20
<210> 382
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> downstream amplification primer for SEQ 185, SEQ 262, SEQ 186, SEQ 263,
SEQ 187, SEQ 264
<400> 382
gactgtatcc tttgatgcac 20
<210> 383
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind

<222> 1..20
<223> downstream amplification primer for SEQ 188, SEQ 265, SEQ 189, SEQ 266
<400> 383
gcataattgt gcttgactgg 20
<210> 384
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer for SEQ 190, SEQ 267, SEQ 191, SEQ 268
<400> 384
tgctgagagg agcttttg 18
<210> 385
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer for SEQ 192, SEQ 269, SEQ 193, SEQ 270
<400> 385
tgaggactgc taggaaag 18
<210> 386
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> downstream amplification primer for SEQ 194, SEQ 271
<400> 386
acaaaatcag gaacaatggg 20
<210> 387
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer for SEQ 195, SEQ 272, SEQ 196, SEQ 273
<400> 387
ttgcattttc cccccaac 18
<210> 388
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer for SEQ 197, SEQ 274, SEQ 198, SEQ 275,
SEQ 199, SEQ 276
<400> 388
accatttgga caatgggg 18
<210> 389
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind

<222> 1..20
<223> downstream amplification primer for SEQ 200, SEQ 277, SEQ 201, SEQ 278
<400> 389
gctcttaaagc tggctctgtg 20
<210> 390
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer for SEQ 202, SEQ 279, SEQ 203, SEQ 280
<400> 390
ggcatgactt cacgtttc 18
<210> 391
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer for SEQ 204, SEQ 281, SEQ 205, SEQ 282,
SEQ 206, SEQ 283, SEQ 207, SEQ 284, SEQ 208, SEQ 285
<400> 391
aggatcttct acagtcac 18
<210> 392
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> downstream amplification primer for SEQ 209, SEQ 286, SEQ 210, SEQ 287
<400> 392
tggtagcgtt tgaaatcatc 20
<210> 393
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> downstream amplification primer for SEQ 211, SEQ 288
<400> 393
tataagcaca aataggttcc 20
<210> 394
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer for SEQ 212, SEQ 289, SEQ 213, SEQ 290
<400> 394
gaataactga ggggagtg 18
<210> 395
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind

<222> 1..19
<223> downstream amplification primer for SEQ 214, SEQ 291, SEQ 215, SEQ 292
<400> 395
gtgaatctcc ttccaag 19
<210> 396
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer for SEQ 216, SEQ 293, SEQ 217, SEQ 294
<400> 396
ctaagtggt gtagacag 18
<210> 397
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> downstream amplification primer for SEQ 218, SEQ 295, SEQ 219, SEQ 296
<400> 397
cacctcgata aatcaagtc 20
<210> 398
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> downstream amplification primer for SEQ 220, SEQ 297, SEQ 221, SEQ 298
<400> 398
gttcacttaa ttctgttgag 20
<210> 399
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer for SEQ 222, SEQ 299, SEQ 223, SEQ 300
<400> 399
cgcttttct gaaaggtg 18
<210> 400
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer for SEQ 224, SEQ 301, SEQ 225, SEQ 302
<400> 400
atcttctgca cagcagcg 18
<210> 401
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19

<223> downstream amplification primer for SEQ 226, SEQ 303
<400> 401
tattttctag ctcttctgg 19
<210> 402
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> downstream amplification primer for SEQ 227, SEQ 304
<400> 402
agcaagagtg attgtaaag 19
<210> 403
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer for SEQ 228, SEQ 305
<400> 403
tattcagaaa ggagtggg 18
<210> 404
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer for SEQ 229, SEQ 306, SEQ 230, SEQ 307,
SEQ 231, SEQ 308, SEQ 232, SEQ 309
<400> 404
agagcgttct tgcctttc 18
<210> 405
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> downstream amplification primer for SEQ 233, SEQ 310
<400> 405
ggtaacccta aaatgttatc 20
<210> 406
<211> 21
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..21
<223> downstream amplification primer for SEQ 234, SEQ 311
<400> 406
agaaaccata agggatatatt g 21
<210> 407
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19

<223> downstream amplification primer for SEQ 235, SEQ 312
<400> 407
acagtgc aaa gggtatatc 19
<210> 408
<211> 21
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..21
<223> downstream amplification primer for SEQ 236, SEQ 313
<400> 408
gaacaacctt gaattagctt g 21
<210> 409
<211> 21
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..21
<223> downstream amplification primer for SEQ 237, SEQ 314
<400> 409
gattccagaa gtccatttca g 21
<210> 410
<211> 21
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..21
<223> downstream amplification primer for SEQ 238, SEQ 315
<400> 410
aggtaagaat gagcaaaaag g 21
<210> 411
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> downstream amplification primer for SEQ 239, SEQ 316, SEQ 240, SEQ 317
<400> 411
gcttggtgtt gttcaattc 19
<210> 412
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer for SEQ 241, SEQ 318, SEQ 242, SEQ 319
<400> 412
cttgaaatac tcccagcc 18
<210> 413
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19

<223> downstream amplification primer for SEQ 243, SEQ 320, SEQ 244, SEQ 321, SEQ 245, SEQ 322, SEQ 246, SEQ 323, SEQ 247, SEQ 324, SEQ 248, SEQ 325, SEQ 249, SEQ 326 -
<400> 413
ccatgaactg agaactttg 19
<210> 414
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer for SEQ 250, SEQ 327
<400> 414
ggtgacaggt aaagaaac 18
<210> 415
<211> 21
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..21
<223> downstream amplification primer for SEQ 251, SEQ 328
<400> 415
attcaggcac agaagtcata c 21
<210> 416
<211> 21
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..21
<223> downstream amplification primer for SEQ 252, SEQ 329, SEQ 253, SEQ 330
<400> 416
agggcagcac aatgtagtaa g 21
<210> 417
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer for SEQ 254, SEQ 331
<400> 417
cctctttatc tccaaacc 18
<210> 418
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> downstream amplification primer for SEQ 255, SEQ 332
<400> 418
gaaaacaatc aagctctgg 19
<210> 419
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind

<222> 1..19
<223> downstream amplification primer for SEQ 256, SEQ 333
<400> 419
cctttatatc cttggagtc 19
<210> 420
<211> 21
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..21
<223> downstream amplification primer for SEQ 257, SEQ 334
<400> 420
tattacacgt tccaactctt c 21
<210> 421
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> downstream amplification primer for SEQ 258, SEQ 335
<400> 421
ctgtgtttaa gtgactgctg 20
<210> 422
<211> 21
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..21
<223> downstream amplification primer for SEQ 259, SEQ 336
<400> 422
ttattgcccc acatgcttga g 21
<210> 423
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> downstream amplification primer for SEQ 260, SEQ 337
<400> 423
tcattcgtct ggctaggtc 19
<210> 424
<211> 21
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..21
<223> downstream amplification primer for SEQ 261, SEQ 338
<400> 424
gaaacagact gaagcaagga c 21
<210> 425
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19

<223> potential microsequencing oligo for 4-14-107.misl
<400> 425
acaaccacca aatgcatac 19
<210> 426
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-14-317.misl
<400> 426
acatgcaagg tgggcaaga 19
<210> 427
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-14-35.misl
<400> 427
aacacagaaa ccgctaaaa 19
<210> 428
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-20-149.misl
<400> 428
tttttgcgtgt gtcttcaaag tga 23
<210> 429
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-20-77.misl
<400> 429
acatgaagat tctgaaggg 19
<210> 430
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-22-174.misl
<400> 430
ggattgtgca gaagttgcct ttc 23
<210> 431
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-22-176.misl

<400> 431
tgcagaagtt gcctttcat 19
<210> 432
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-26-60.mis1
<400> 432
ggaaagtgca tcttaagac 19
<210> 433
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-26-72.mis1
<400> 433
ttaagacagt tagcaggcc 19
<210> 434
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-3-130.mis1
<400> 434
gggcctaaaa cagtattct 19
<210> 435
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-38-63.mis1
<400> 435
agttataaga aaatcaggc 19
<210> 436
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-38-83.mis1
<400> 436
gaggctaaac ttttttttt 19
<210> 437
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-4-152.mis1
<400> 437

```
ttccattgt tcctgactt 19
<210> 438
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-4-187.mis1
<400> 438
tataaacaga aacatggatg agt 23
<210> 439
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-4-288.mis1
<400> 439
catcaactaa ttttcacaa 19
<210> 440
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-42-304.mis1
<400> 440
tttaaaacta tttatgtaa 19
<210> 441
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-42-401.mis1
<400> 441
taagaaagaa ttctgtgttc tgg 23
<210> 442
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-43-328.mis1
<400> 442
ttctgtgttc tggccaaag 19
<210> 443
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-43-70.mis1
<400> 443
atcgctcca ttatttcaa aaa 23
```

<210> 444
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-50-209.mis1
<400> 444
atatagagtg tgcattccctg aca 23
<210> 445
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-50-293.mis1
<400> 445
cctgagtccc agggggctga cag 23
<210> 446
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-50-323.mis1
<400> 446
ttttaaacaat tgatgaatct tta 23
<210> 447
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-50-329.mis1
<400> 447
acattgatga atctttatta cta 23
<210> 448
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-50-330.mis1
<400> 448
gatgaatctt tattactac 19
<210> 449
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-52-163.mis1
<400> 449
gaacaggata tttttaacta cca 23
<210> 450

<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-52-88.mis1
<400> 450
tccatgtcat tattattcaa aag 23
<210> 451
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-53-258.mis1
<400> 451
aatcatgcag agagaatgc 19
<210> 452
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-54-283.mis1
<400> 452
aagtagtttt tcacactttc tct 23
<210> 453
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-54-388.mis1
<400> 453
ctatcgtata catctttac 19
<210> 454
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-55-70.mis1
<400> 454
aagaacctag gttttaaaa 19
<210> 455
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-55-95.mis1
<400> 455
ctctctatcg tatacatctt tac 23
<210> 456
<211> 23

<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-56-159.mis1
<400> 456
aagttttcct tctcttctgt aga 23
<210> 457
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-56-213.mis1
<400> 457
ctcatgttca ctctgggtc 19
<210> 458
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-58-289.mis1
<400> 458
catacctgca gcctgctttt ggt 23
<210> 459
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-58-318.mis1
<400> 459
tgactacttt acctgcaata ttt 23
<210> 460
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-60-266.mis1
<400> 460
aacaggacca agacactgca tta 23
<210> 461
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-60-293.mis1
<400> 461
aagtttcagt atttcttagc aga 23
<210> 462
<211> 19
<212> DNA

<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-84-241.misl
<400> 462
aaaaaatagt gactgccac 19
<210> 463
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-84-262.misl
<400> 463
tgaataattc agttcttca 19
<210> 464
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-86-206.misl
<400> 464
tcaaattcagg acacaccac 19
<210> 465
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-86-309.misl
<400> 465
tctaggcagg ccactttag 19
<210> 466
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-88-349.misl
<400> 466
ctaaaagaca atattcagt 19
<210> 467
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-89-87.misl
<400> 467
ttcttcctg aacgctggtt tca 23
<210> 468
<211> 19
<212> DNA
<213> Homo Sapiens

<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-123-184.mis1
<400> 468
cccagaacat tcaccagct 19
<210> 469
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-128-202.mis1
<400> 469
tctgtttctt agagaactg 19
<210> 470
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-128-275.mis1
<400> 470
ccctacctca catgtgtag 19
<210> 471
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-128-313.mis1
<400> 471
tctctagaca gatatacat 19
<210> 472
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 99-128-60.mis1
<400> 472
cactgtgacc caggcgctag cgt 23
<210> 473
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-12907-295.mis1
<400> 473
tatggcatta tatctccac 19
<210> 474
<211> 19
<212> DNA
<213> Homo Sapiens
<220>

<221> primer_bind
<222> 1..19
<223> microsequencing oligo for 99-130-58.misl
<400> 474
caaaagagct tcaaaaata 19
<210> 475
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-134-362.misl
<400> 475
acactcatgt tagttagat 19
<210> 476
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> microsequencing oligo for 99-140-130.misl
<400> 476
caaaagcagc tacagacca 19
<210> 477
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> microsequencing oligo for 99-1462-238.misl
<400> 477
ttcaagggtta gtaactcat 19
<210> 478
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-147-181.misl
<400> 478
catgaaaaag agcatgata 19
<210> 479
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-1474-156.misl
<400> 479
tactcataag ttaaattatt 19
<210> 480
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind

<222> 1..19
<223> potential microsequencing oligo for 99-1474-359.misl
<400> 480
aaaatcaaatt tattgtacc 19
<210> 481
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> microsequencing oligo for 99-1479-158.misl
<400> 481
aaaatccact tgtaatcgc 19
<210> 482
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-1479-379.misl
<400> 482
agctgtgtac tgaggtcag 19
<210> 483
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-148-129.misl
<400> 483
tatctataca aataatttt 19
<210> 484
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-148-132.misl
<400> 484
ctatacaaat aattttgaa 19
<210> 485
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-148-139.misl
<400> 485
aataattttg aatttaata 19
<210> 486
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19

<223> potential microsequencing oligo for 99-148-140.misl
<400> 486
ataattttga atttaatac 19
<210> 487
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-148-182.misl
<400> 487
tgttgataatg ggcaactgt 19
<210> 488
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-148-366.misl
<400> 488
tgtcaaaggt ctctccctg 19
<210> 489
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-148-76.misl
<400> 489
agaatgcctt cctgaatta 19
<210> 490
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-1480-290.misl
<400> 490
ccatcttcac cacaacccc 19
<210> 491
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-1481-285.misl
<400> 491
ataacctgtt ttgcttctc 19
<210> 492
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-1484-101.misl

<400> 492
agatcaaata taagcatgt 19
<210> 493
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> microsequencing oligo for 99-1484-328.mis1
<400> 493
acgtggtcat gaggagttt 19
<210> 494
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-1485-251.mis1
<400> 494
gccttgatat atgctccca 19
<210> 495
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> microsequencing oligo for 99-1490-381.mis1
<400> 495
cagtggaaat accatgtca 19
<210> 496
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-1493-280.mis1
<400> 496
gacagagtat tgttgagg 19
<210> 497
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-151-94.mis1
<400> 497
agatcattga taaggaaat 19
<210> 498
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> microsequencing oligo for 99-211-291.mis1
<400> 498

ttatatcaga ctgaccttc 19
<210> 499
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-213-37.mis1
<400> 499
cttccggctg caggactgt 19
<210> 500
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-221-442.mis1
<400> 500
tttgtagata tgcattggga 19
<210> 501
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 99-222-109.mis1
<400> 501
caggtgagga gtgctggatt ggc 23
<210> 502
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-14-107.mis2
<400> 502
ctatcaggca ttgctctgtg tgc 23
<210> 503
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-14-317.mis2
<400> 503
tcattgacctg tgccacctc ttt 23
<210> 504
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-14-35.mis2
<400> 504
tctctgcaga cagcttctgc ctg 23

<210> 505
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-20-149.mis2
<400> 505
agcaggcaat aaaccaaga 19
<210> 506
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-20-77.mis2
<400> 506
gtgttctcag acaacaaag 19
<210> 507
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-22-174.mis2
<400> 507
aaattaacat ttttgaaca 19
<210> 508
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-22-176.mis2
<400> 508
acaaattaac atttttgaa 19
<210> 509
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-26-60.mis2
<400> 509
aagtcgctcc tcggcctgct aac 23
<210> 510
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-26-72.mis2
<400> 510
accctttaaa gtcgctcct 19
<210> 511

<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-3-130.mis2
<400> 511
agttaataacc aatttaagct tta 23
<210> 512
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-38-63.mis2
<400> 512
aaaaaaaaag tttagcctc 19
<210> 513
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-38-83.mis2
<400> 513
atattctcaa cagcattgcc aaa 23
<210> 514
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-4-152.mis2
<400> 514
tgtttatata taggataac 19
<210> 515
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-4-187.mis2
<400> 515
tttttttttt ttttttttt 19
<210> 516
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-4-288.mis2
<400> 516
tgaaatcaaa acataggta 19
<210> 517
<211> 19

<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-42-304.mis2
<400> 517
aaaaaccct gaaaataag 19
<210> 518
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-42-401.mis2
<400> 518
tctgtgggtt taaactttg 19
<210> 519
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-43-328.mis2
<400> 519
actggctctg tgggtttaa 19
<210> 520
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-43-70.mis2
<400> 520
ttgtgttggtg tcccatggt 19
<210> 521
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-50-209.mis2
<400> 521
cataaagcct tcagtttca 19
<210> 522
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-50-293.mis2
<400> 522
tcaatgtttt aaactgtcc 19
<210> 523
<211> 19
<212> DNA

<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-50-323.mis2
<400> 523
atcgaaccct tttgtagta 19
<210> 524
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-50-329.mis2
<400> 524
gcctaaatcg aaccctttt 19
<210> 525
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-50-330.mis2
<400> 525
agcctaaatc gaacccttt 19
<210> 526
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-52-163.mis2
<400> 526
aatagatgtg taaaattct 19
<210> 527
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-52-88.mis2
<400> 527
caccttgtgt attttttaa 19
<210> 528
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-53-258.mis2
<400> 528
ttaggttaaa atttgagtga gaa 23
<210> 529
<211> 19
<212> DNA
<213> Homo Sapiens

<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-54-283.mis2
<400> 529
taagccatcg attgtatca 19
<210> 530
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-54-388.mis2
<400> 530
gtcttggcgc tgcagcgtg 19
<210> 531
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-55-70.mis2
<400> 531
taaagatgta tacgatagag agt 23
<210> 532
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-55-95.mis2
<400> 532
gtcttggcgc tgcagcgtg 19
<210> 533
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-56-159.mis2
<400> 533
ttgactgtaa catggagac 19
<210> 534
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-56-213.mis2
<400> 534
tatcaaactc ctctgaagg 19
<210> 535
<211> 19
<212> DNA
<213> Homo Sapiens
<220>

<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-58-289.mis2
<400> 535
aggtaaagta gtcaccct 19
<210> 536
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-58-318.mis2
<400> 536
gaagaaataa acttgcaaa 19
<210> 537
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-60-266.mis2
<400> 537
agaaatactg aaactttat 19
<210> 538
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-60-293.mis2
<400> 538
aggacttcct gctggcttc 19
<210> 539
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-84-241.mis2
<400> 539
ttctgaagaa ctgaattatt cac 23
<210> 540
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-84-262.mis2
<400> 540
tgagatcatg ttgctgctt 19
<210> 541
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind

<222> 1..23
<223> microsequencing oligo for 4-86-206.mis2
<400> 541
aatgttaacg tgtagatgcc att 23
<210> 542
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-86-309.mis2
<400> 542
gctctctggg tcttcactc 19
<210> 543
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-88-349.mis2
<400> 543
aagaacttgg aaaatctca 19
<210> 544
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-89-87.mis2
<400> 544
ttctcaacac aaaaactat 19
<210> 545
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-123-184.mis2
<400> 545
cccagcagaa ctcttggcc 19
<210> 546
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-128-202.mis2
<400> 546
gtatgtatgt gtgtgtgtt 19
<210> 547
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19

<223> potential microsequencing oligo for 99-128-275.mis2
<400> 547
acatatatgc atacatttg 19
<210> 548
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-128-313.mis2
<400> 548
tctatgccaa atagaatct 19
<210> 549
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-128-60.mis2
<400> 549
ggagtgtcac tgtaagagg 19
<210> 550
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> microsequencing oligo for 99-12907-295.mis2
<400> 550
ttgtacatca ggtctgccc 19
<210> 551
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 99-130-58.mis2
<400> 551
ctcgccatat gcacactcct gaa 23
<210> 552
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> microsequencing oligo for 99-134-362.mis2
<400> 552
tctttgtaat aggaataat 19
<210> 553
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> microsequencing oligo for 99-140-130.mis2

<400> 553
catgctcaat tgtttacat 19
<210> 554
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-1462-238.mis2
<400> 554
ctgaagcaga aacacagca 19
<210> 555
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 99-147-181.mis2
<400> 555
tataaagatt taagtttttc ttt 23
<210> 556
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> microsequencing oligo for 99-1474-156.mis2
<400> 556
ccatatttct tcttggtat 19
<210> 557
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-1474-359.mis2
<400> 557
catctgatat tagggaatt 19
<210> 558
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-1479-158.mis2
<400> 558
aatatacact ccaattagc 19
<210> 559
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-1479-379.mis2
<400> 559

```
ctgtaccatg agctgcttc 19
<210> 560
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-148-129.mis2
<400> 560
cagccctatg tattaaatt 19
<210> 561
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-148-132.mis2
<400> 561
ttgcagccct atgtattaa 19
<210> 562
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-148-139.mis2
<400> 562
ccttggtttg cagccctat 19
<210> 563
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-148-140.mis2
<400> 563
accttggttt gcagcccta 19
<210> 564
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 99-148-182.mis2
<400> 564
gaatgctttg ggaccatcca aca 23
<210> 565
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-148-366.mis2
<400> 565
gaggcggcag ccgtgagca 19
```


<210> 566
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-148-76.mis2
<400> 566
tatgaagcca tcaagagta 19
<210> 567
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> microsequencing oligo for 99-1480-290.mis2
<400> 567
aaaaggatca gtggttgcc 19
<210> 568
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> microsequencing oligo for 99-1481-285.mis2
<400> 568
taccatcttg aggttagag 19
<210> 569
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-1484-101.mis2
<400> 569
agattttaag gagaggagt 19
<210> 570
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-1484-328.mis2
<400> 570
tctgaaaact gaatccctt 19
<210> 571
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> microsequencing oligo for 99-1485-251.mis2
<400> 571
aggggacatt cttggttct 19
<210> 572

<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-1490-381.mis2
<400> 572
agatgcacag tagcgtacc 19
<210> 573
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> microsequencing oligo for 99-1493-280.mis2
<400> 573
acaagcagcc aaaccccat 19
<210> 574
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 99-151-94.mis2
<400> 574
atatagattt tgaaatttta gaa 23
<210> 575
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-211-291.mis2
<400> 575
cattgacctg ttgaaaaca 19
<210> 576
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-213-37.mis2
<400> 576
tcagacactg gagtcctcc 19
<210> 577
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-221-442.mis2
<400> 577
gtctggctag gtcattggaa 19
<210> 578
<211> 19

<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-222-109.mis2
<400> 578
ctgaagaaat tcatatcgt